The Development and Testing of Risk Adjusters Using Medicare Inpatient and Ambulatory Data

FINAL REPORT

Prepared for:

Health Care Financing Administration
Contract No. 93-026/EE

Prepared by:

The Lewin Group
Allen Dobson, Ph.D.
Kevin Coleman

The Johns Hopkins University
Jonathan Weiner, Dr. P.H.
Gerard Anderson, Ph.D.
Barbara Starfield, M.D., M.P.H.
Stephanie Maxwell

August 5, 1996

ACKNOWLEDGMENTS

The primary authors of this report are Allen Dobson, Ph.D., Project Director, with assistance of Kevin Coleman, Project Coordinator of The Lewin Group; and Jonathan Weiner, Dr. P.H., Principal Investigator; Gerard Anderson, Ph.D., and Barbara Starfield, M.D., Co-Investigators and Stephanie Maxwell, Project Manager of the Johns Hopkins University.

The authors would especially like to recognize the energy and effort of the programming staff of Johns Hopkins University: Yifei Hu, Programming Analysis; Chad Abrams; and Andrew Baker.

The authors gratefully acknowledge the support and assistance of Mel Ingber, Ph.D., the Project Officer for this contract. His insights and feedback throughout the study were particularly helpful in identifying and correcting technical concerns and in the important lessons and policy implications of our work.

Table of Contents

Table of Contents (Continued)

Table of Contents (Continued)

LIST OF TABLES

LIST OF APPENDICES

EXECUTIVE SUMMARY

The Health Care Financing Administration (HCFA) is continuing to support research to develop risk adjuster models that combine demographic and clinical data to predict the expenditures of Medicare enrollees in order that payments to capitated plans might better reflect underlying enrollee disease burden. This project is one of two that HCFA has supported that developed and then evaluated new risk adjuster models for the Medicare population using newly available ambulatory diagnosis codes found within Medicare administrative claims data. For this project, researchers at Johns Hopkins University (JHU) developed two different risk adjuster models that their colleagues at The Lewin Group then evaluated. A team of researchers from Boston University (BU) and the Center for Health Economics Research (CHER) conduct a second, parallel project to design and evaluate other risk adjuster models for the Medicare population.

The organization of this Executive Summary and the accompanying report echoes the organizational structure of the Lewin/JHU project. The two primary activities of this project -- model development and model evaluation -- were divided between The Lewin Group and JHU. JHU designed the two risk adjuster models for this project, while The Lewin Group conducted the evaluation of these models. This organization of tasks minimized interaction between model development and model evaluation. The Executive Summary opens with a brief Background and Overview section. Next, the development of the two new JHU risk adjuster models designed for this project is discussed. These two models are then evaluated from two different perspectives -- their predictive accuracy; and the feasibility of using these models as the foundation of a new capitated payment system for Medicare enrollees. The Executive Summary ends with a conclusions section.

A. BACKGROUND AND OVERVIEW

Before proceeding with an overview of our project, we first discuss the conceptual differences among "risk assessment," "risk adjustment," and "risk adjusted capitated payments systems." Using risk assessment and risk adjustment methods as the foundation for a capitated payment system is not new. HCFA currently uses a risk assessment/adjustment model, the Adjusted Average per Capita Cost (AAPCC) model, to reimburse managed care organizations (MCOs) that enroll Medicare beneficiaries. The limitations of the AAPCC model, however, provide both a context and a rationale for the development of more powerful risk adjuster models.

1. Risk Assessment, Risk Adjustment, and Risk Adjusted Capitated Payment Systems

Differences in the use of medical services across individuals are to some degree predictable. The use of medical services depends in part on observable characteristics, including demographic and clinical factors as well as the prior use of medical services. "Risk assessment" models use these demographic, clinical, and prior use data to classify individuals according to their expected use of medical services as compared to other members of an insured population. "Risk adjustment" uses the results of "risk assessment" to convert individual differences in the expected use of medical services into dollar premium values -- typically per member per month payments -- for a given year.

The purpose of risk adjustment is not to predict all medical expenditures, but rather only those that can be predicted using available data and measurement technology. Most of the variation in expenditures cannot be explained, which provides a role for health insurance. The task of risk adjustment is to improve the functioning of health insurance by measuring and then paying for predictable health care expenditures, thereby limiting or preventing selection in the health insurance marketplace. Ideally, risk adjustment would transfer revenues from insurers experiencing favorable selection to their counterparts who experience unfavorable selection. By controlling for selection effects across insurers, risk adjustment reduces or eliminates the incentives for plans to seek out the healthy enrollee, while avoiding the sick.

There are many uses of risk assessment and risk adjustment. For example, insurers and health care organizations often use risk assessment and adjustment measures as part of their clinical profiling activities. That is, the variations in clinical practice among providers can be compared after a risk adjustment model helps to control for predictable differences in medical service utilization across individual patients.

Risk assessment and risk adjustment may also be used when designing capitated payment systems. Risk adjustment, however, is only one part of the overall design of a risk adjusted capitated payment system. Third party payers designing such a system also face a series of technical and policy decisions that are not directly addressed by a risk assessment or risk adjustment system. For example, how payment amounts are updated to reflect changes in the price and composition of medical services and whether the payment system includes adjustments for differences in costs across regions are two such policy issues. In addition, risk adjustment information can be augmented with reinsurance or diagnostic carve-outs. In this Executive Summary and in the full report we discuss other key decisions a third party payer will face when designing a risk adjusted capitated payment system.

This project refined and extended several different existing risk assessment technologies. These risk assessment technologies were then used to create two risk adjustment models that predicted individual differences in medical service expenditures in 1992 for a large sample of Medicare beneficiaries based on the prior year's information. Either of these two risk adjustment models could serve as one of the fundamental building blocks of a risk adjusted capitated payment system.

2. The AAPCC

Medicare now uses the AAPCC system to set capitated payments for risk contract MCOs for Medicare enrollees. The AAPCC uses age, gender, Medicaid status, and institutional (nursing home) status to create a series of mutually exclusive rate cells. Starting in 1995, the AAPCC also incorporated working-aged rate cells. These rate cells are designed to reflect the costs of providing care to Medicare enrollees treated in a fee-for-service setting. HCFA sets capitated payments at 95 percent of costs predicted by the AAPCC rate cells.

The AAPCC has been criticized on numerous grounds. The criticisms include: (1) the AAPCC's lack of predictive accuracy for each individual enrollee; (2) a questioning of the methods used to calculate the payment amounts associated with each cell as well as the adjusters for differences in local prices and practice pattern variation; (3) the use of fee-for-service costs to predict costs in MCO settings; (4) possible negative incentives that encourage MCOs to enroll "marginal" institutional patients and to keep patients in low-cost institutional settings; (5) the use of geographic adjustments that may perpetuate inefficiencies in service delivery; and (6) the potential for increasing divergence of payments based on fee-for-service costs relative to MCO costs in areas where MCO market share is growing.

B. MODEL DEVELOPMENT

The JHU team members were responsible for developing both risk adjuster models for this project. Model development activities included: (1) extending the Ambulatory Diagnostic Group (ADG) risk assessment system designed by JHU mainly for the under-65 population to the Medicare over-65 population; (2) updating the Payment Amount for Capitated Systems (PACS) risk adjuster model that was also developed by JHU; and (3) developing one or more combined risk adjuster models using the updated ADG and PACS risk assessment/adjustment technologies for the Medicare over 65 population. This section first discusses the ADG and PACS model components, then the development of the database used in model development and later in model evaluation, and finally describes each of the two JHU risk adjuster models.

The models developed for this project did not use medical expenditures in year one to help predict medical expenditures in year two. While many previous studies (Epstein and Cumella, 1988) have demonstrated that prior medical expenditures are a strong predictor of future medical expenditures, prior medical expenditures were not considered an appropriate component of a risk adjustment model by the study team for two reasons. First, prior medical expenditures are highly susceptible to gaming by plans and providers. Second, incorporating prior expenditures into a risk adjuster model begins to approximate retrospective, cost-based payment that is well-known to be highly inflationary.

1. Model Components

The two JHU risk adjuster models use demographic and diagnostic information from year one to predict medical expenditures in year two. The JHU team began model development for this project by refining and updating two risk assessment technologies previously developed by JHU researchers. The first of these risk assessment technologies is based upon Ambulatory Diagnostic Groups (ADGs). [1] ADGs classify nearly all ICD-9-CM diagnoses into one of 34 distinct diagnostic categories, each containing numerous ICD-9-CM codes. The ICD to ADG grouping process is based on whether: (1) the diagnosis will persist or recur; (2) return visits and/or continued treatment will be needed; (3) specialist services will be required; (4) the life expectancy of individuals with a given diagnosis decreases; (5) the diagnosis will result in short or long-term disability; (6) needed diagnostic and therapeutic procedures will be expensive; and (7) hospitalization will be required.

The ADGs were originally developed for a primarily under 65 working-age population. A portion of the development work was to extend the ADG concept to the over 65 population. For a fixed time period (usually one year), an individual may have diagnoses that place him or her into any number of the 34 different ADGs. ADGs as part of the ACG system, are now commonly used for clinical profiling, research, and other uses by MCOs as well as other health care organizations. Based on development activities, thirteen of the 34 ADG categories were identified as being most predictive of future resource use among the elderly and were adopted for inclusion in our final models.

[1] ADGs are the building blocks of the "Ambulatory Care Group" (ACG) methodology, an ambulatory case-mix classification system now in use at over 100 organizations. The ACG system collapses the large number of possible ADG patient combinations into 52 mutually exclusive ACG categories, including a "no-ADG" category. The development team explored the use of the mutually exclusive ACGs, as opposed to the ADG system, in its risk adjusters. The ADGs proved to be more predictive of year two medical expenditures for the elderly; thus, they were used in the final models.

The second existing risk assessment/risk adjustment technology that provided the basis for our new method is Payment Amounts for Capitated Systems (PACS). The development of PACS, an inpatient oriented risk adjustment system, was sponsored by HCFA as a potential prospective payment methodology for risk-contract health maintenance organizations (HMOs). The original PACS combined three demographic characteristics (age, gender, and disability status) with three claims-based measures: (1) the major diagnostic category (MDC) (of the DRG system) associated with each hospital admission of a patient in the base year; (2) the chronicity of each primary clinical diagnosis; [2] and (3) whether the Part B Medicare deductible was met (a measure of ambulatory care use).

As discussed below, both JHU models included the three demographic variables used by the PACS model (along with a Medicaid eligibility variable as well). The Medicare Part B deductible was dropped from model development, because ADGs were used to measure ambulatory care use levels. The PACS chronicity measure was used to help reassign some ICD-9-CM codes across ADG groups, but was not otherwise incorporated into our final model.

This left the most important component of PACS, the MDCs. The MDCs were selected from a range of inpatient utilization classification systems. Developed as part of the diagnostic related group (DRG) system, MDCs group all inpatient admissions into one of 27 broad organ-system categories based on the patients' primary hospital discharge ICD-9-CM diagnosis.

2. Database Development

This project made use of HCFA's Standard Analytic Files (SAFs) created from the National Claims History File. SAFs contain 100 percent of the institutional bills and a five percent sample of the physician/supplier claims. Since 1991, the physician/supplier claims include ICD-9-CM diagnosis codes for ambulatory visits on almost all records.

The development of the analytic database used for this project involved a high degree of collaboration between the Lewin/JHU and the BU/CHER research teams. In a parallel project, the BU/CHER team developed a revised version of the Diagnostic Cost Group (DCG) risk assessment/adjustment system using the same database.

The project data were drawn from five HCFA SAFs from 1991 and 1992. These included the five percent national, random sample of physician/supplier, inpatient, outpatient, home health, and hospice claims. These claims data were merged with demographic data on Medicare beneficiaries from HCFA's Health Insurance Skeleton Eligibility Write-off (HISKEW) file. The standard five percent sample consisted of approximately 1.5 million aged (age 65 and above) Medicare beneficiaries.

Several groups were excluded from the final analytic database. These included beneficiaries: (1) enrolled in HMOs; (2) lacking Part B coverage; (3) below the age of 65 who were eligible for Medicare disability insurance (DI) or the End Stage Renal Disease (ESRD) program; and (4) not continuously enrolled for 1991 and 1992 (those who died in 1991 were also excluded, but those who died in 1992 were included in the final analytic database). After these restrictions were applied, the resulting final analytic database totaled 1.24 million individuals. The final analytic database was then split into two half-samples, a "development" half-sample used for model development, and a "test" half-sample used for model evaluation.

[2] Chronicity indicates how likely an individual with an inpatient admission in a MDC in a base year is to have additional inpatient admissions in the future. For example, an inpatient admission for trauma has low chronicity, while an inpatient admission related to a chronic condition (e.g., asthma) has higher chronicity.

3. Designing the JHU Risk Adjuster Models

The over-arching goal of this project was to develop a practical method of risk assessment/adjustment that could be used as part of a risk adjusted capitated payment system for Medicare beneficiaries. This project's basic premise is that demographic, clinical, and limited prior use data can be used to predict an individual's use of medical resources in some future period. After considering many different models, the development team designed its risk adjustment/assessment models using the analytic tools and risk assessment technologies described below.

a. Rate Cells vs. Regression Analysis

The main purpose of the risk adjuster models developed during this project is to use diagnostic data from a base year (1991 or "year one") to predict annual medical expenditures in a future year (1992 or "year two"). Some risk assessment/adjuster models use base year data to divide the population into a series of mutually exclusive groups, or actuarial cells. Then, each individual in an insured population can be assigned to one of these cells. Ideally, the expected medical expenditures of individuals within each of these cells in year two should be similar, and the groups themselves should be clinically cogent.[3]

These actuarial rate cells thus constitute a risk assessment system. Risk adjustment occurs when each individual in a given group is assigned a value equal to the average expenditures of members of that cell in year two. For example, the 52 ACGs could serve as a risk adjuster model with 52 rate cells. The expenditures assigned to each risk cell in models of this type are then converted to capitated payments that are made to plans.

JHU designed the risk adjuster models for this project in a different manner. JHU used year one data not to assign individuals to a rate cell, but instead to assign each individual a risk score. An individual's risk score depends on the unique combination of his or her risk assessor characteristics. Using these risk scores, the JHU models then predict each individual's annualized expenditures for year two. Finally, a capitated payment system based on such a risk adjustment model would use the predicted expenditures for each individual to set capitation payments for an enrollee group.

For this study, JHU developed two new risk adjuster models, an "ADG-MDC" and an "ADG-Hosdom" model. Both models include the same demographic variables as well as the 13 ADG variables used to represent diagnostic codes found in the claims data. The MDC model uses a set of 15 MDC variables to represent inpatient admissions, while the ADG-Hosdom model represents inpatient care using the Hosdom variable (described below). One final difference between the two models is how each defines ADGs. The ADG-MDC model defines ADGs using only ambulatory diagnostic information because of the presence of the MDC variables. In contrast, the ADG-Hosdom model defines ADGs using both ambulatory and inpatient diagnostic data. We describe the building blocks of the ADG-MDC and ADG-Hosdom models below in more detail.

[3] It is possible that the base period and future period, or payment period, may not be a year in length. It is also possible that the lag between the base period and payment period is not equal to a year.

b. Expenditures in Year Two: The Dependent Variable

The dependent (outcome) variable in each regression equation for the two risk adjusters is Medicare's annual expenditures for each sample beneficiary in year two (1992). the Lewin/JHU and BU/CHER project teams worked with HCFA-ORD to determine a common, best method for specifying annual Medicare expenditures in 1992. Physician/supplier expenditures were estimated using the Resource-Based Relative Value Scale (RBRVS) amounts adjusted by the Geographic Practice Cost Index (GPCI) to reflect local prices. Claims payment amounts were used for the less common physician/supplier services where RBRVS units have not been assigned. Inpatient expenditures were estimated using hospital specific DRG payments that included capital, outlier, medical education and wage index payments. Expenditures for most other services were estimated using claims payment amounts. Non-Medicare services (e.g., outpatient drugs) were of course not included. Finally, Medicare copayments and deductibles and claims where Medicare was the secondary payer were excluded.

c. Risk Assessors: The Independent Variables

Each JHU model then used a series of risk assessors as the independent (explanatory) variables in the risk adjuster regression equations. The coefficients for each variable indicate the "payment weight" associated with that variable in 1992 -- i.e., the expected increase in 1992 medical expenditures associated with that specific risk assessor. These "payment weights," which reflect annual payment amounts linked to each risk factor, are presented in ***Table ES-1.*** Each individual's risk score is the sum of payment weights for his or her risk assessment characteristics. The risk assessors included in each regression equation are described further below.

i. The ADG-MDC Model

Both JHU risk adjuster models share the same set of demographic risk assessor variables. The intercept term in each model indicates the expected 1992 expenditures of a "base case" individual. This person is female, age 65, with no ADG diagnoses or MDC inpatient admissions in 1991, and who has never been eligible for Medicare as a result of receiving Social Security Disability Insurance benefits ("ever disabled") and who was not eligible for Medicaid benefits in any month during 1991 (Medicaid). The ADG-MDC model predicts that this individual's annual Medicare expenditures in 1992 would have been $608. All risk adjuster expenditure estimates are made relative to this base case. For example, the ADG-MDC model increases the expenditure estimates by $604 for men, by $67 for each year of age over age 65, by $1,119 for someone who was never disabled, and by $761 for someone who was Medicaid eligible.

The two JHU models differ according to their clinical and prior use risk assessors. Both models include a set of 13 ADG "dummy" variables (yes/no) indicating if an individual had a diagnosis that was grouped into that ADG in 1991. For example, the MDC-ADC-risk adjustment score of an individual with visits for a single diagnosis in ADG 3 in 1991 (Time limited diagnosis, major) would increase by $542 in 1992 ***(Table ES-1)***, but would not increase for any other additional diagnoses that may also have been categorized in that ADG. This individual's risk score, however, would increase if he or she had a diagnosis that fell into one of the other ADGs included in the ADG-MDC model in 1991. The ADGs in the ADG-MDC model are based on diagnoses assigned by providers during face-to-face encounters in an ambulatory setting.

Table ES-1
JHU Risk Adjuster Annual Weights: 1992

Variables	ADG-MDC		ADG-Hosdom	
	Weight	Std. Error	Weight	Std. Error
Demographic Variables				
Intercept	$608	$28	$434	$28
Male	604	26	613	26
Years over Age 65	67	2	64	2
Ever Disabled	1,119	51	1,176	52
Medicaid	761	43	802	43
Hospital Dominant Marker				
Hosdom			1,749	43
MDCs				
MDC 1 (Nervous System)	1,533	36		
MDC 3 or 4 (Ear, Nose, Throat and Respiratory)	3,237	46		
MDC 5 (Circulatory System)	1,879	79		
MDC 6 (Digestive System)	1,759	30		
MDC 7 (Hepatobilliary System and Pancreas)	1,030	53		
MDC 8 (Musculoskeletal System and Connective Tissue)	1,117	27		
MDC 9 (Skin, Subcutaneous Tissue and Breast)	1,762	77		
MDC 10 (Endocrine, Nutritional, and Metabolic Diseases)	2,938	43		
MDC 11 (Kidney and Urinary Tract)	2,526	116		
MDC 16, 17, or 25 (Blood, Immunological, Myeloproliferative Diseases, and AIDS/HIV)	3,061	79		
MDC 18 (Infectious and Parasitic Diseases)	1,957	32		
MDC 19 or 20 (Mental Diseases and Alcohol/Drug Abuse)	1,882	29		
MDC 21 or 22 (Injuries, Poisonings, and Burns)	1,481	40		
MDC 23 or 24 (Factors Influencing Health Status and Trauma)	3,875	79		
MDC 26 (Transplants)	3,944	60		
ADGs		-		
ADG 3 (Time Limited, Major)	542	36	663	35
ADG 4 (Time Limited, Major, Primary Infections)	734	64	1,503	44
ADG 6 (Asthma)	818	123	1,216	76
ADG 7 (Likely to Recur, Discrete)	225	65	365	30
ADG 9 (Likely to Recur, Progressive)	965	134	1,696	49
ADG 11 (Chronic Medical, Unstable)	1,345	126	1,415	27
ADG 16 (Chronic Specialty, Unstable, Orthopedic)	650	107	593	74
ADG 22 (Injuries/Adverse Effects, Major)	525	177	462	40
ADG 23 (Psychiatric, Time Limited, Minor)	698	110	1,222	107
ADG 25 (Psychiatric, Persistent or Recurrent, Unstable)	804	245	1,088	69
ADG 27 (Signs/Symptoms, Uncertain)	460	163	568	32
ADG 28 (Signs/Symptoms, Major)	551	97	753	30
ADG 32 (Malignancy)	1,347	206	1,429	40

Note: Each payment weight reflects the additional annual capitation payment for year two associated with the presence of the risk assessment characteristics in year one. The MDC payment weights are based on each year-one admission. The ADG and Hosdom payment weights are based on one or more service contacts in year one.

Neither model used all 34 ADGs that comprise the ADG system. Several ADGs were eliminated because aged Medicare beneficiaries were unlikely to have diagnoses in these ADG categories (e.g., pregnancy). Other ADGs were eliminated after it was determined that no statistically significant relationship existed between these ADGs and annual expenditures in 1992.

The remaining risk assessors in the ADG-MDC risk adjuster model are the MDCs. These MDCs indicate the number of inpatient admissions each individual had in 1991 in each major diagnostic category. In contrast with ADGs, MDCs are count variables -- an individual's risk score increases for each inpatient admission that individual had in the base year (1991) in that MDC. For example, an individual's risk score in 1992 would increase by $1,533 for each inpatient admission that person had in MDC 1 (the nervous system) in 1991. The ADG-MDC model uses only 15 of the 26 MDCs from the PACS model. Some MDC categories were combined into a single category in the ADG-MDC model (e.g., MDCs 22 and 23). Finally, other MDCs were not included, again because aged Medicare beneficiaries were unlikely to have inpatient admissions in these categories or because these MDCs for 1991 were not statistically significantly associated with Medicare expenditures in 1992.

ii. The ADG-Hosdom Model

The ADG-Hosdom model uses the same 13 ADGs as the ADG-MDC model, as well as the same demographic variables. In the ADG-Hosdom model, however, ADGs are defined by diagnoses made during face-to-face provider contacts in both ambulatory and inpatient settings, as indicated on HCFA claims data[4].

In addition, the ADG-Hosdom model does not include MDC variables. Instead, a new risk assessor variable, the "hospital-dominant" ("Hosdom") variable, was defined. A diagnosis (ICD-9-CM code) in 1991 is considered a "Hosdom" diagnosis if at least 50 percent of Medicare beneficiaries with that diagnosis had an inpatient admission for that diagnosis during 1991. The Hosdom risk measure was developed as a marker for serious conditions, that usually lead to hospitalization given current patterns of practice. Unlike the MDC-based measures, however, to be categorized as having the Hosdom marker does not require hospitalization, and thus would not penalize a provider for choosing ambulatory-based treatment. The ADG-Hosdom model increases the risk adjuster expenditure score of individuals with one or more Hosdom diagnoses in 1991 by $1,749.

C. EVALUATING THE JHU RISK ADJUSTER MODELS

After developing the models, the JHU team delivered them to The Lewin Group team along with the split-half sample database to be used for model evaluation. The major focus of The Lewin Group's evaluation of the two JHU risk adjuster models was to assess their predictive accuracy. This section first discusses the key findings of The Lewin Group's assessment of the JHU risk adjuster models' predictive accuracy. In addition, The Lewin Group team members also assessed the potential that either JHU model could be "gamed" by insurance plans and providers, as well as considering the administrative feasibility of each model. This assessment of gaming and administrative feasibility concludes this section.

[4] Face-to-face diagnoses are made using the procedure codes (HCFA Common Procedure Coding System (HCPCS)) on the claim. Face-to-face encounters are defined as visits involving an evaluation and/or management service or procedure performed by a physician (MD or DO) or a limited license professional (nurse practitioner, physician's assistant, dentist, podiatrist, social worker, chiropractor, or psychologist).

1. Evaluation Measures

The Lewin Group evaluation team principally used two measures to compare the predictive accuracy of the ADG-MDC and ADG-Hosdom models to each other and to an AAPCC model. First, each model's ability to account for individual variation in medical expenditures was measured using adjusted R square statistics for different groups. The adjusted R square statistics were calculated by reestimating each risk adjuster model for each non-random group from the evaluation split-half sample. Risk adjuster models with greater individual predictive accuracy are more resistant to "cherry picking" by plans, because it would be more difficult for plans to identify the "best" risks and because the financial rewards for selecting the best risks would be smaller.

Second, the ability of each model to predict the expenditures of entire groups of Medicare beneficiaries was assessed using predictive ratios. Predictive ratios are the expected expenditures for a particular enrollee group as predicted by a given risk adjuster model divided by the actual expenditures of that group. A predictive ratio of 1.00 indicates that the risk adjuster model predicted the average expenditures of a group perfectly. In contrast, predictive ratios of less than 1.00 indicate the risk adjuster model under-predicted the expenditures of the group in question, while predictive ratios of greater than 1.00 indicate over-prediction. Risk adjusters tend to under-predict the expenditures of enrollee groups with higher severity of illness and over predict expenditures for low severity groups.

Adjusted R square statistics and predictive ratios were also calculated for a version of the AAPCC model. This model included gender, age over 65 (as a continuous variable), Medicaid buy-in and ever having been disabled.[5] The inclusion of results for the "AAPCC" provides a comparative context for interpreting findings for the ADG-MDC and ADG-Hosdom models.

2. Random Groups

The first series of evaluation "enrollee" groups considered in this analysis were repeated random groups of different sizes. One hundred groups of 500, 1,000, 5,000, 10,000, and 50,000 individuals were selected at random from the test split half sample. For each set of 100 groups, predictive ratios were calculated; adjusted R square statistics for these randomly selected groups were not calculated. We present these results in ***Table ES-2*** for groups of 500, 5,000, and 50,000 individuals. These results indicate that as expected, the distribution of predictive ratios for all three risk adjustment models became more closely centered around 1.00 (median values) as the size of these random groups increased.

The results for the random groups of 50,000 individuals indicate there is a small chance for favorable or negative risk selection for both JHU models due to random chance alone. That is, both JHU models under-predicted or over-predicted the expenditures of approximately five percent (i.e., the 5th and 95th percentiles) of these groups by approximately 10 percent. On the other hand, both JHU models generally predicted the medical expenditures of these large random groups somewhat more accurately than did the "AAPCC." This was true, even though the Law of Large Numbers would imply that demographic risk adjuster models such as the "AAPCC" should predict the expenditures of large, random groups quite accurately.

[5] Throughout the remainder of this Executive Summary and report, when referring to the version of the AAPCC model we estimated, we surround the AAPCC acronym in quotation marks. When we are referring to the actual AAPCC model used by HCFA or the AAPCC as an abstract concept, we use the AAPCC acronym without quotation marks.

Table ES-2
Distribution of Predictive Ratios for 100 Random Samples of Different Sizes

Model	5th Percentile	25th Percentile	Median	75th Percentile	95th Percentile
Group Size: 500					
"AAPCC"	0.8115	0.9073	1.0367	1.1140	1.3352
ADG-MDC	0.8355	0.9092	1.0312	1.1047	1.2968
ADG-Hosdom	0.8304	0.9169	1.0370	1.0977	1.2879
Group Size: 5,000					
"AAPCC"	0.8590	0.9175	0.9776	1.0521	1.1698
ADG-MDC	0.8602	0.9297	0.9839	1.0593	1.1477
ADG-Hosdom	0.8683	0.9391	0.9892	1.0582	1.1444
Group Size: 50,000					
"AAPCC"	0.8901	0.9040	1.0040	1.0461	1.1325
ADG-MDC	0.9063	0.9344	0.9972	1.0545	1.1127
ADG-Hosdom	0.9129	0.9363	1.0002	1.0410	1.1139

It is possible that some of the risk of over-prediction or under-prediction was due the presence or absence of an unusual number of very high cost individuals (i.e., individuals with medical expenditures in 1992 of $50,000 or $100,000 or more). To test this possibility expenditures were truncated at these two thresholds ($50,000 and $100,000) and each of the three risk adjuster models was re-estimated. These truncated models were then combined with stop-loss reinsurance of 80 percent above these two thresholds. Finally, the predictive ratios for the random groups of 50,000 individuals were recalculated for the ADG-MDC, ADG-Hosdom, and "AAPCC" risk adjuster models under these two reinsurance schemes. As described in the body of this report, the effects of reinsurance on the predictive ratios for random groups of 50,000 were quite small. Reinsurance, however, may still play an important role if it provides protection against catastrophic loss for plans with relatively few enrollees.

3. Non-Random Groups

The second type of enrollee groups used to test the risk adjuster models are non-random groups. Testing the performance of risk adjuster models on non-random groups provides additional information on their relative performance. For instance, a given risk adjuster model may be particularly well-suited to predict the expenditures of a given group defined by age and gender but less able to predict the expenditures for the remaining age/gender groups.

The non-random groups in this analysis were selected after consultation with our project officer and other HCFA staff, and included the following groups defined using year one (1991) individual characteristics:

- **Age/gender** -- where the age groups are 65 to 69, 70 to 74, 75 to 79, 80 to 84, and 85+;
- **Race** -- white, black, and other;
- **Hospital admissions** -- one, two, and three or more hospital admissions;

- **Heavy users of ambulatory services** -- these are individuals with no hospital visits but whose use of physician services is high (i.e., their use of physician relative value units (RVUs) is in excess of one standard deviation above the mean number RVUs) in a single year;
- **Medical conditions** -- enrollees with one or more of the following 17 medical conditions: (1) depression; (2) alcohol and drug abuse; (3-4) hypertension; (5-6) diabetes; (7-9) cardiac conditions; (10) pulmonary conditions; (11-13) cancers; (14-15) stroke; (16) hip fracture; and (17) arthritis;
- **Expenditure groups** -- quintiles of annualized payment -- the first quintile group has the lowest annual medical expenditures in 1991, while the fifth quintile group had the highest; and
- **Geographic location** -- the nine census division regions.

These non-random groups each provide information on the relative ability of each risk adjuster model to limit HMO selection bias. The performance of each risk adjuster model across these groups indicates how well a given model might operate for groups differentiated by age, gender, and race; for groups with low or high use of medical services and/or medical expenditures; and for groups in different regions of the county. It is important to note, however, that these groups represent extreme cases -- i.e., what happens if HMOs enrolled only individuals with a certain medical condition or those who were especially low or high users of medical services.

This evaluation clearly demonstrated the superiority of the JHU risk adjuster models relative to the "AAPCC." For the entire split-half sample, the adjusted R square statistics for the ADG-MDC model (6.3 percent) and the ADG-Hosdom model (5.5 percent) were five to six times as large as those of the "AAPCC" model (1.0 percent).[6] For virtually all of the tested non-random groups, the adjusted R square statistics for both JHU risk adjuster models were much higher than for the "AAPCC," indicating that the JHU models were better able to account for individual variation in medical expenditures for the selected non-random groups.

In addition, the predictive ratios for the JHU risk adjuster models consistently clustered more tightly around 1.00 for the non-random groups than did the predictive ratios for the "AAPCC" model. Again, this indicates the superior performance of the JHU risk adjusters, in this case for predicting the medical expenditures of non-random groups.

The predictive ratio results were not always consistent with the adjusted R square statistics in measuring the performance of risk adjuster models. In most cases, the ADG-MDC model had slightly higher adjusted R square statistics than did the ADG-Hosdom model for most non-random groups. This is due likely to the inclusion of the MDC variables in the ADG-MDC model, which can better capture the expenditures of individuals whose use of inpatient medical services is greater. There was no correspondence, however, between the adjusted R square statistics and the predictive ratios of the ADG-MDC and ADG-Hosdom models. That is, the ADG-Hosdom model frequently out-performed the ADG-MDC model for some non-random groups by having predictive ratios that were nearer to 1.00, even though the adjusted R square

[6] The adjusted R-square statistics reported here for the ADG-HDC, ADG-Hosdom, and "AAPCC" models were for the test split-half sample. The corresponding adjusted R-square statistics for the development split-half sample were slightly higher for the ADG-MDC (6.3 versus 6.0 percent) and ADG-Hosdom models (5.5 versus 3.3 percent)

statistics for these non-random groups were lower for the ADG-Hosdom than for the ADG-MDC model.

One of the most important strengths of both JHU models relative to the AAPCC is their potential ability to reduce the gains from cream-skimming. This point is displayed graphically in ***Chart ES-1***. Here, the predictive ratios for the ADG-MDC, ADG-Hosdom, and "AAPCC" models are displayed for groups defined by their medical service expenditures in 1991. For instance, the "First Quintile" indicates those Medicare enrollees who were in the bottom fifth of medical service expenditures in 1991. For this group, the "AAPCC" on average would pay plans 234 percent of the group's actual medical expenditures in 1992 or 2.34 times their actual costs. Thus, the AAPCC provides plans with strong incentives to identify and enroll individuals with low costs in the past.

In contrast, the ADG-MDC model only over-predicts the expenditures of this least expensive group by 19 percent in 1992, while the ADG-Hosdom model over-predicts the expenditures of this enrollee group by only eight percent in 1992. Thus, the incentives for cream-skimming under either JHU risk adjuster model are greatly reduced. The "AAPCC" under-predicts the medical expenditures in 1992 of the group in the top fifth of medical service use in 1991 by 50 percent (the "fifth" quintile). Here, the AAPCC provides plans with strong incentives to disenroll or avoid enrolling individuals with a history of high medical service use. Again, both JHU models do a much better job of predicting the expenditures of this high cost group -- the ADG-Hosdom model under-predicts the expenditures of this group by 12 percent, while the ADG-MDC model under-predicts these expenditures by only eight percent. This would imply that the gains to plans from cream-skimming would be sharply reduced were capitated payments based on either JHU risk adjuster model.

From the perspective of both HCFA and the plans, a model's ability to predict the expenditures of individuals with particular chronic conditions is important (***Table ES-3***). If a model consistently under-predicts the expenditures of individuals with chronic conditions, these individuals may have problems with access to care, and plans that could care for these patients, even efficient plans, would be discouraged from doing so. Conversely, if a model over-predicts the expenditures of individuals with a given chronic condition, plans would have strong financial incentives to enroll these individuals.

Both the ADG-MDC and ADG-Hosdom models do a far better job of predicting the medical expenditures in year two (1992) of groups with the chronic conditions presented in ***Table ES-3*** than does the "AAPCC" model. For example, the ranges of predictive ratios across the 17 chronic condition categories are narrower for the ADG-MDC (0.72 to 1.42, or a range of 0.70) and ADG-Hosdom models (0.66 to 1.42, or a range of 0.76) than the "AAPCC" model (0.34 to 1.35, or a range of 1.01). In only two cases, diabetes without complications and breast cancer, does the "AAPCC" model have predictive ratios nearer to 1.00 than the two JHU risk adjuster models. In 10 cases, the ADG-MDC model has predictive ratios closer to 1.00 than the other models, while the ADG-Hosdom model had the "best" predictive ratios for the other five groups.

CHART ES-1
PREDICTIVE RATIOS FOR 1991 EXPENDITURE QUINTILE GROUPS

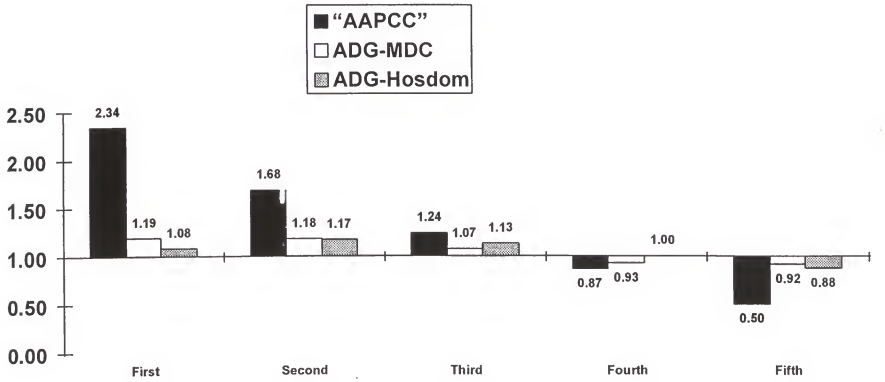


Table ES-3
Adjusted R Square Statistics and Predictive Ratios by Year One (1991) Diagnostic Categories

Diagnosis	"AAPCC"		ADG-MDC		ADG-Hosdom	
	Adjusted R. Square a/	Ratio b/	Adjusted R. Square a/	Ratio b/	Adjusted R. Square a/	Ratio b/
Depression	0.76%	0.9437	0.45%	0.9921	4.69%	1.0215
Alcohol /Drug Abuse	-0.13%	0.7918	-6.12%	1.1128	-3.20%	1.2096
Hypertensive Heart/Renal Disease	0.77%	1.1712	3.54%	1.1664	2.57%	1.2091
Benign/Unspecified Hypertension	0.62%	1.3546	2.51%	1.0564	2.23%	1.0643
Diabetes with Complications	0.96%	0.8854	3.83%	1.0301	2.65%	1.0591
Diabetes without Complications	0.63%	0.9260	3.44%	0.8528	3.03%	0.8621
Heart Failure/Cardiomyopathy	0.17%	0.7133	3.80%	0.8965	3.16%	0.8810
Acute Myocardial Infarction	0.09%	0.6335	1.92%	0.8827	1.98%	1.0071
Other Heart Disease	0.53%	0.7873	3.37%	1.0353	2.78%	1.0354
Chronic Obstructive Pulmonary Disease	0.86%	0.6834	5.83%	0.9415	4.59%	0.9238
Colorectal Cancer	0.30%	0.5383	5.42%	0.8734	3.95%	0.8981
Breast Cancer	0.54%	0.9270	6.22%	1.4189	5.16%	1.4223
Lung/Pancreas Cancer	3.10%	0.3360	4.93%	0.7150	3.97%	0.3589
Other Stroke	0.49%	0.5638	4.91%	0.9355	4.19%	0.9911
Intracerebral Hemorrhage	-0.53%	0.4415	-1.54%	0.8111	-0.11%	0.9203
Hip Fracture	0.16%	0.6525	3.63%	0.9704	2.68%	1.0531
Arthritis	0.80%	0.8151	5.15%	0.9572	4.57%	0.9773

Note: The adjusted R-square statistics in this table were calculated by re-estimating each model for every group in the test split-half sample. The predictive ratios were calculated by applying the payment weights for each model from the entire development split-half sample.

a/ The adjusted R square statistic can be negative in certain circumstances. These circumstances include regression models that include a large number of explanatory variables and that explain little of the variation in the dependent variable (here, medical expenditures in year two), especially when such a model is estimated using a data set with relatively few observations .

b/ The predictive ratio represents the group's expected expenditure (i.e., the capitation rate) divided by the group's actual expenditure for year-two.

All three risk adjuster models either over-predict or under-predict the medical expenditures of given chronic condition groups. For example, all three models over-predict the expenditures of both hypertension groups. Thus, plans that enroll individuals with hypertension might be rewarded under a capitated payment system based on any of these three risk adjuster systems. Conversely, plans could incur losses if they enroll individuals with diabetes without complications, heart failure/cardiomyopathy, chronic obstructive pulmonary disease, colorectal cancer, intracerebral hemorrhage, or arthritis, if capitated payments depended on the ADG-MDC, ADG-Hosdom, or "AAPCC" risk adjuster systems. In most cases, however, the two JHU models would lead to less severe levels of under payment.

4. Gaming and Administrative Feasibility

One potential problem with the diagnosis-based JHU models or other diagnosis-based risk adjuster models is the potential for "upcoding" by plans and providers. This would occur if plans engaged in strategic behaviors to increase risk adjusted scores by their enrollees, such as by recording additional diagnoses or reclassifying diagnoses. Under the AAPCC, it is unlikely that plans can upcode, because payment depends on gender, age, Medicaid status, and institutional status -- non-diagnostic variables that are easily verified and mostly coded by HCFA in any event.[7] It is possible that the ADG-Hosdom model could be particularly susceptible to upcoding, given that plans would need to code only a single Hosdom diagnosis in one year to receive a large payment increase in the next year.

Several factors reduce the ability of plans to game either the ADG-MDC or ADG-Hosdom models through upcoding. First, it will take time for plans to identify the best options for gaming, and yet more time for plans to acquire the data collection and manipulation skills needed for successful upcoding. Second, if all plans engage in some level of upcoding, the payer (HCFA) could respond to this "code creep" by reducing the rate of increase in overall payment levels to maintain budget neutrality. Another choice would be to recalibrate ("rebase") the model periodically to control for changes in the prevalence of "code creep." Third, the payer could adopt simple auditing procedures to identify obvious examples of gaming.

Another issue is whether plans can exploit informational advantages to risk select against a risk adjuster. Newhouse et al. (1989) and others argue that plans will have access to data, such as prior expenditure data, that better predict the future medical expenditures of enrollees than the demographic and clinical variables used by risk adjuster models. If so, plans could identify and target the best risks for enrollment under any risk adjuster. As noted earlier, we do not advise using prior medical expenditures as a risk assessor, because of the susceptibility of this measure to gaming by plans and providers and because of the inherent inflationary effects of basing payment on prior medical expenditures that is roughly a form of retrospective cost-based reimbursement.

In addition, it appears clear that both JHU models provide much less scope for plans to "cream-skim" than do less sophisticated risk adjusters, such as the AAPCC. Thus, the marginal benefits may not exceed the marginal costs of "cream skimming" if payments are based on a sophisticated, diagnosis-based risk adjuster system. This would tend to reduce the incentives to cream-skim for plans. Finally, as experience with more powerful methods of risk adjustment is gained, payment systems will become more sophisticated and will develop better defenses to gaming. This could include incorporating many of the data elements available to

[7] The AAPCC also includes rate cells for enrollees with disabilities and for enrollees with end-stage renal disease (ESRD). It is conceivable, however, that plans could encourage the enrollment of a disproportionately large or small number of plan members with disabilities or ESRD.

plans in the risk adjuster model directly. Ideally, this will force plans to compete on the basis of price and quality, not through gaming and risk selection.

One potential problem for the ADG-MDC model is that it could encourage inappropriate hospital admissions, since the ADG-MDC model increases payments in year two for each hospital admission in year one. These additional payments, however, are much less than the costs of these additional admissions. On the other hand, there may be cases where these payments would exceed the difference in costs to plans of caring for patients in non-inpatient settings and admitting these patients to a hospital. If so, this could provide some incentive for plans to increase admissions under the ADG-MDC model.

There are also a number of other administrative issues that must be addressed before either JHU model could serve as the basis of a risk adjusted, capitated payment system. The risk adjuster models developed for this project used data from 1991 and 1992 to predict medical expenditures, or payment weights, in 1992. To use these models for payment purposes, HCFA will need to select some method of updating their payment weights to the current year. In addition, the payment weights for the two JHU models are national payment weights. Given that the current AAPCC methodology HCFA now uses adjusts payments at the county level to reflect differences in the costs of providing care across different plan areas, it is also likely that some method of converting the JHU model's national payment weights to county payment amounts must also be designed.

One way of updating payment weights would be to use the percentage change in the U.S. per capita cost (USPCC) or some other factor to update these 1992 payment weights. There are at least two limitations of this approach. First, the relative payment weights are likely to change as medical practice and medical technology changes. Repeatedly updating an initial set of payment weights cannot account for these relative changes. Thus, HCFA may want to rebase its relative national payment weights by reestimating the regression equations underlying the ADG-MDC and ADG-Hosdom models using the most recently available data. In addition, the payment system may also be updated to reflect any ongoing developments in ADG, MDC, and Hosdom coding.

These two limitations could be solved in the following way. Suppose HCFA will be using either JHU risk adjuster model to set county-specific payment amounts in 1997. The first step would be to reestimate the model using individual demographic and diagnostic data from 1994 to estimate individual medical expenditures in 1995. These are the most recent years for which these data would be available for 1997 payment calculations. This would yield an updated set of national payment weights for 1995. These weights could then be adjusted for inflation to 1997, perhaps by using USPCC inflation rates.

In the second step, HCFA would compile 1996 demographic, diagnostic and cost data for each county. Using the cost data from 1996 for Medicare fee-for-service (FFS) enrollees for each county, HCFA would compute a county per capita cost (CPCC) amount.[8] These CPCCs would then be adjusted for inflation to convert them in 1997 amounts. Note that these CPCCs include no adjustments for any demographic or diagnostic factor. The CPCC in each county will indicate to plans the capitated payment amount they can expect to receive for an "average" enrollee in that county.

In the third step, HCFA would estimate a FFS average payment weight (FAPW) using one of the JHU models (that have been reestimated using 1994 and 1995 data) and the

[8] HCFA may wish to make these calculations using moving average of several years of county FFS cost data, to smooth out any unusual yearly fluctuations.

demographic and diagnostic data for FFS enrollees from 1996 in each county. HCFA would use the risk adjuster model to estimate a risk score for each FFS enrollee (this risk score would be equal to the sum of that individual's payment weights for each risk assessment factor). The county FAPW would then equal the sum of the risk scores for all FFS enrollees in that county divided by the total number of county FFS enrollees.

In the final step, HCFA would use the CPCCs and FAPWs to calculate payment amounts for risk contract enrollees in each county using the following formula:

$$\frac{Risk\ Score_{ij}}{FAPW_j} * CPCC_j$$

where i is an enrollee in a risk contract in county j. The payment amount for enrollee i thus equals enrollee i's risk score divided by the county FAPW and then multiplied by the county CPCC.

Another administrative issue that must be addressed is dealing with new Medicare enrollees and with individuals enrolled for less than a year. New Medicare enrollees will lack the diagnostic data required to assign these individuals risk adjustment scores. Some interim payment must be adopted to pay for these new enrollees until HCFA has sufficient diagnostic data to calculate their risk adjustment scores. The issue of new plan enrollees that were previously Medicare eligible are easier to address, as Medicare data can be used to calculate their risk adjustment scores. If enrollees are allowed to change plans frequently as is likely with Medicare, it is likely that some enrollees will be enrolled in a plan for less than a year. Tracking these partial year enrollees and keeping their payment status updated could pose problems.

It is also likely that risk contractors will take time to adjust to the new data requirements associated with using a diagnostic risk adjuster to set capitated payments. While ICD-9-CM codes must now be filed on all ambulatory and inpatient Medicare claims, the accuracy and completeness of these diagnostic data likely will improve and perhaps become more timely after plan payments begin to depend on these data.

There are also other administrative details that must be worked out. Will diagnostic risk adjusters be adopted abruptly, or will they be phased-in? Will other risk adjustment mechanisms, such as reinsurance or diagnostic carve-outs, be integrated with a diagnostic-based risk adjuster system? How long will it take plans to become familiar and comfortable with diagnostic-based risk adjuster systems, especially regression-based systems? Finally, will national rates be used, or will geographic and regional adjustments be added, as they now are in the AAPCC? Each of these questions must be answered before a diagnosis-based risk adjusted, capitated payment system could be adopted by HCFA.

D. CONCLUSIONS

Both JHU risk adjuster models are clear improvements over the current AAPCC. The ADG-MDC and ADG-Hosdom models have much greater predictive accuracy than does our version of the "AAPCC" model, both at the individual and group level and for both randomly selected and non-randomly selected groups of the population. By accounting for more of the predictable variation in medical expenditures, both JHU models limit the potential for "gaming" by plans and providers. The results for non-random groups are especially important, indicating that if capitated payments were based on either JHU model rather than the AAPCC, the financial rewards associated with plans' "cream-skimming" of Medicare beneficiaries would be greatly reduced. By incorporating diagnostic and clinical data, the JHU risk adjuster models are

far more powerful and have much greater clinical acceptance and cogency than does the AAPCC.

There are some limitations of the JHU models. As with any payment system based on risk adjustment, the JHU models would provide plans and providers with opportunities for upcoding and other gaming activities. The project team believes, however, that such gaming would be costly and difficult, and detection of serious gaming or outright fraud should not be that arduous or expensive for payers that adopt either model as the basis of a capitated payment system.

Work remains to be done before either JHU model can be used to set capitated payments. Any new risk adjusted payment proposal would require demonstration before it could be implemented as part of the Medicare payment formula for MCOs. In particular, the current Medicare payment and delivery systems demonstrations offer excellent opportunities for further study of both JHU risk adjusters from many perspectives -- statistical, clinical, and administrative.

There are also several areas for future research that are worth pursuing. First, further development of the Hosdom variable could reduce the potential of upcoding by providers. For example, instead of one Hosdom diagnosis, the variable could be reconfigured so that two diagnoses or an inpatient admission is needed to trigger the Hosdom variable. In addition, the ADG-MDC and ADG-Hosdom models could be extended to include condition specific per case "carve-out" payments and/or reinsurance. In another project, the Lewin/JHU team are designing and evaluating carve-outs and reinsurance as part of developing new risk adjuster models for an under age 65 population for HCFA.

An important final point to make is that while the development of risk assessment/adjustment systems is time-consuming and technically complex, the use of these systems for payment purposes would be straight-forward. Once a system of payment rules has been designed, tested, and implemented, these rules then could be incorporated into a stand-alone software package similar to the DRG "grouper" and PPS "pricer" programs. While these software programs for diagnostic risk adjusters would be "black boxes," it would not be difficult for HCFA to develop or run them either centrally or locally. Once tested and implemented, these programs would create risk adjusted per member per month payments for all Medicare beneficiaries enrolled in capitated plans.

In conclusion, there is reason for optimism that risk adjusted payments can be made powerful enough to support a more level playing field in MCO competition for Medicare beneficiaries. If so, competition based on premium price and quality would be encouraged and under service for sicker patients and cream-skimming would be diminished.

CHAPTER I
INTRODUCTION

A. BACKGROUND

In this introduction, we first define risk adjustment. We then overview the current system of risk adjustment used by Medicare and note its conceptual and practical shortcomings. This is followed by a summary of project goals and objectives. We then conclude the introduction with a discussion of the clinical underpinnings of the risk adjustment models developed by this project.

1. Risk Adjustment Defined

In the context of health insurance, a "risk adjuster" is defined as a method for classifying individuals according to their expected health care resource use. Usually, a risk adjuster is based on a set of clinical, demographic or prior health care utilization factors. Risk adjusters can be used to explain current or past resource use, or to predict future resource consumption of groups of individuals. A major application of risk adjusters is to establish capitation payments to more accurately reflect the degree of risk associated with providing services to a specific "enrolled" population.

Ideally, a risk adjustment system used for payment purposes should incorporate a reliable measure of the health status of individuals, which is the primary determinant of health service utilization. Measuring a person's true health status level is a difficult proposition, from both a conceptual and empirical perspective. This difficulty is magnified further when the goal is to develop a rating for all members within a very large insured population such as the Medicare program. For these reasons, factors that can be readily measured for a population – such as age, gender, previous encounters with the delivery system, and diagnoses assigned by providers – are typically used as proxies for health status. Generally, persons with similar utilization, morbidity or demographic characteristics can be expected to use similar levels of health care resources over a given period of time.

2. Current Medicare HMC Risk Adjusted Payment Method: AAPCC

The Social Security Act Amendments of 1972 authorized both cost and risk reimbursement on behalf of Medicare beneficiaries to HMOs. The option of risk reimbursement introduced the need for a risk based payment system. The "adjusted average per capita costs" (AAPCC) method of payment was developed in response to this need. Section 1876 of the 1972 Amendments identified risk adjustments to assure "actuarial equivalence including adjustments related to age distribution, sex, race, institutional status, disability status, and any other relevant factors."

Section 1876 was later revised by the Tax Equity and Fiscal Responsibility Act (TEFRA) of 1982. TEFRA Section 114, implemented in 1985, required that:

> *"The Secretary shall annually determine a per capita rate of payment for each class of individuals entering into risk contracts ... the Secretary shall define the appropriate classes of members based on age, disability status, and other such factors as the Secretary determines to be appropriate ... the annual per capita rate of payments for each class shall be equal to 95 percent of the adjusted average per capita cost ..."*

According to the revised language, the Secretary is required to consider age and disability status. Other factors such as institutional status, sex, and welfare/Medicaid are not specifically required in the revised legislation.

The Office of the Actuary of the Health Care Financing Administration (HCFA) annually calculates an estimate of the AAPCC value for the prospective service year for each HMO. The process of calculating the AAPCC requires three steps:

- HCFA calculates the United States Per Capita Cost (USPCC), which is the projected Medicare expenses for the average Medicare beneficiary in the next year. Using actual and historical Medicare claims data, six USPCCs are calculated: Part A for the aged, the disabled, and people with end stage renal disease, and Part B for the same three groups;
- County-level geographic adjustments are applied to the USPCC; and
- HMO-specific "demographic" adjustments for age, gender, Medicaid status, and institutional status of enrollees are applied to the geographically adjusted USPCC. A further adjustment to the USPCC for enrollee work status was added in 1995.

HMOs electing payment under a Medicare risk contract receive payment at a per capita rate of 95 percent of the AAPCC for each Medicare enrollee in their plan.

a. Criticisms Of AAPCC

The AAPCC has been criticized on technical and conceptual grounds. First, a large body of literature has demonstrated that the adjustment factors currently used (age, sex, Medicaid status and institution status) are not accurate predictors of an enrollee's future use of resources (Anderson, Resnick and Gertman 83; Hornbrook 84; Beebe, Lubitz, and Eggers 85). The most common statistical measure of risk adjuster accuracy is the "adjusted R-square" statistic. In terms of risk adjustment, this statistic measures the percent of variance in actual health expenditures that is explained by a risk adjuster payment model. The AAPCC method explains 0.3 to 0.6 percent of the variance in annual Medicare covered expenditures on an individual beneficiary level (Lubitz, Beebe, and Riley 85; Thomas, Lichetenstein, Wyszewianski, et al., 83 Thomas, Berki, Liechtenstein et al. 85).

Although most agree that 0.6 percent is low, it is not clear what a satisfactory percentage would be. Health expenditures largely are random, and thus unforeseen by either the individual or the HMO. Joseph Newhouse has demonstrated that the approximate maximum percentage of variation in health care expenses that could be explained is 12 to 18 percent (Newhouse et al. 89).

A second set of criticisms have focused on the methods of calculation for various adjustment factors relating to local price and practice pattern variation.

The third criticism attacks the underlying assumption on which the AAPCC rests: that HMO payments should be based on 95 percent of the adjusted average per capita cost incurred by beneficiaries participating in the FFS sector. To the extent that the health status of beneficiaries differ between the HMO and FFS sectors, or that practice patterns differ, the FFS average may not serve as an appropriate benchmark.

A fourth problem is that certain AAPCC factors create incentives for "upcoding." For example, by paying HMOs more if the beneficiary comes from an institutional setting, a financial incentive is created to enroll beneficiaries from marginal institutional settings such as residential facilities, and to "upcode" their living arrangements to "institutional" for additional payment purposes.

A fifth issue with the current method pertains to the geographic adjustment. A geographic adjustment likely perpetuates regional differences in health care costs that are based on inefficient

or inequitable medical practices in some areas. By basing the AAPCC on the FFS costs in the local area, HCFA is capturing all of the inefficiencies of the FFS plus the geographic variation in utilization rates.

A sixth potential concern is that over time the AAPCC formula may not be the most appropriate method of estimating premiums. As the percent of HMO-enrolled Medicare beneficiaries in a locality grows, the number of patients remaining in the FFS sector will decrease. If skewed selection is occurring in either the HMO or FFS setting, the FFS experience would not reflect the experience of those in the HMO setting, causing payments to be increasingly distorted over time.

3. Current Medicare Policy Context

The US health care delivery system is changing rapidly. Today, over 80 percent of privately insured Americans under the age of 65 are in some type of managed care plan. Common among these plans is that costs are being aggressively managed, often by using budgeted, non-FFS financing schemes such as capitation or risk sharing with providers. Approximately 30 percent of working aged insured groups are in capitated HMOs. Currently within the Medicare program, about 11 percent of the 37 million beneficiaries are enrolled in various types of HMOs, and 9 percent of this 37 million are enrolled in fully capitated, risk-contract HMOs. Risk contracting is expected to increase in the Medicare program, largely because of pressures to contain costs, as well as possible legislation to encourage broader provider participation in capitated health care arrangements.

Although the subject is of some controversy, recent studies have suggested that the current AAPCC is in fact overpaying capitated plans. One study has estimated that because HMO enrollees are significantly less ill than those beneficiaries that remain in the FFS Medicare system (which is the basis of capitated plans' AAPCC payments), Medicare is thus overpaying its capitated plans (Brown et. al., 93). Another controversial issue relates to the fact that because of different levels of efficiency and waste in the FFS system across various jurisdictions, the AAPCC -- based on underlying FFS costs -- also varies dramatically across regions. This penalizes the HMOs located in "efficient" regions and rewards those in less efficient regions. For example, in 1995 the monthly AAPCC in the metropolitan Miami area was about $615 per person, while in Minneapolis it was about $363. While some of this variation may be due to differences in health status of the elderly in these two locales, as well as cost of living differences, much (if not most) of this 250 percent differential is believed to be due to difference in underlying FFS practice patterns -- hardly an adequate reason to pay HMOs in the two areas such divergent rates (PPRC 95).

The Administration has proposed a series of innovative, managed care plans all expected to receive a fixed per-capita payment (Vladeck 95). Current events on Capitol Hill portend an even more rapid movement to develop a full array of Medicare managed care plans, most of which will be paid a lump-sum per member per month payment for the care of enrolled populations. The success of these scenarios depends upon the development of a risk adjustment method to assure that health plans are paid fairly for the mix of patients they attract and serve. In addition, without some system that more accurately pays Medicare HMOs, Medicare FFS beneficiaries ultimately could be penalized.

B. PROJECT GOALS AND OBJECTIVES

The goal of this HCFA-sponsored research and development contract was to construct and test a risk adjustment method for use in reimbursing Medicare risk-contract HMOs and other capitated providers.

The main objectives of this project were to:

- Extend the Ambulatory Care Group/Ambulatory Diagnosis Group (ACG/ADG) case mix system to the Medicare over-65 population;
- Update the original components of the Payment Amount for Capitated Systems (PACS) risk adjuster;
- Develop one or more prospective risk adjuster models for the over-65 population that integrate inpatient (Medicare Part A) and ambulatory (Medicare Part B) diagnostic information;
- Assess, validate and critique the model(s) from a statistical and practical administrative perspective; and
- Assess the completeness and utility of diagnostic information associated with Medicare ambulatory claims data.

C. CONCEPTUAL BASIS OF MODELS: ADGs and PACSs

The starting point of our new integrated (ambulatory and hospital) risk adjustment method was the combination of two established risk adjuster systems developed previously at The Johns Hopkins University (JHU). These systems are the Ambulatory Care Groups (ACG) case-mix measure and the Payment Amount for Capitated Systems (PACS) inpatient risk adjuster. These risk measures and their relevant component parts are described briefly. (Readers are referred to published sources for a fuller description of the existing ACGs: Weiner et al. 91 and Starfield et al. 91; and PACS: Anderson et al. 90.)

1. Ambulatory Diagnosis Group Morbidity Clusters

One of two major inputs into the final risk adjuster model developed here was the ADG grouping method of the ACG system. ACGs and its building blocks, ADGs, are a population oriented health status tool developed several years ago by JHU researchers. (See ***Appendices I-1*** and ***I-2*** for the original ADG and ACG groups.) ACGs and ADGs were designed to be a conceptually simple, statistically valid, and clinically relevant measure to predict the need for and use of ambulatory health care services. ACGs and ADGs are based on the premise that a measure of a population's "illness burden" can help predict health care resource consumption. A person is grouped first into one or more ADGs, then into a single ACG, based on diagnoses assigned by providers treating them during a predetermined period of time, such as a year. ACGs and ADGs utilize the ICD-9-CM diagnostic codes available in computerized insurance claims or encounter data systems. ACGs and ADGs predict annual relative service use intensity for a given patient, as opposed to visits or encounters.

ACGs and ADGs originally incorporated only ambulatory diagnoses and were intended mainly as a population oriented case-mix measure for use in analyzing or paying for ambulatory care at the aggregated patient group level. The original method was developed largely as a retrospective or concurrent tool to understand current patterns of use. Today, the ACG/ADG system has been extended to include both inpatient and ambulatory codes, and it is being used to manage, capitate and analyze care provided in many types of settings. Over 100 managed care plans are using ACGs for retrospective profiling activities. It has also become a tool for health services researchers to control for case-mix in their claims-based analyses. We are now aware of 36 academic teams in the US and abroad that are using ACGs for such research.

The basic building block of the ACG system is a method for assigning every common ICD-9-CM diagnosis code to one of 34 "morbidity clusters," or ADGs. Every ICD-9-CM falls into only one ADG, but there are many ICD-9-CMs that fall into the same ADG. A person may be assigned many ADGs, as ADGs are not mutually exclusive. Using a decision-branch tree, ADGs and age and gender are used to categorize each person in a population group into one of 52 mutually exclusive ACG categories. Ultimately, this project incorporated only the ADG clusters, and not the ACG end-groups, as a component of the project's risk adjusters.

In the original ACG/ADG development project, the goal of the ICD-9-CM to ADG assignment process was to cluster together similar conditions based on their expected impact on health services resource consumption. The initial assignment of ICD-9-CM codes to ADGs was done by JHU physicians using utilization data from 160,000 patients (mainly under the age of 65) from five test sites. The assignment criteria which were, and continue to be, used as determinants of anticipated health care resource use are clinical judgments concerning:

- Likelihood of persistence or recurrence of the problem;
- Likelihood of return visits and/or the need for continued treatment;
- Likelihood of the need for specialist services;
- Likelihood of decreased life expectancy;
- Likelihood of short-term or long-term patient disability;
- Expected need and cost of diagnostic and therapeutic procedures; and
- Likelihood of a required hospitalization.

The first step in assigning an ADG to a patient entails identifying all unique ICD-9-CM codes attached to specified claims submitted for the analysis period (typically one year, as was used by this project). A main advantage of the ADG system is that since it was originally developed for ambulatory use, it can categorize the 90 percent of the Medicare population who encounter only the ambulatory care system. Every common ICD-9-CM code denoting a primary or subsidiary diagnosis has been assigned to one of 34 ADGs. As noted above, each diagnosis code is grouped into only one ADG, although many different ICD-9-CM codes comprise each ADG. Therefore, a patient being treated for two distinct diagnoses may fall into two separate ADGs or only one ADG, depending on the type of diagnoses. For example, a patient with both obstructive chronic bronchitis (ICD-9-CM code 491.2) and congestive heart failure (ICD-9-CM code 428.0) will fall into one ADG – Chronic Medical: Unstable (ADG 11). Since the ADG system was designed not to take into account frequency of health system contacts, this patient's ADG 11 assignment is the same assignment as a patient who was diagnosed with, for example, congestive heart failure alone.

The current project's clinical team assigned several hundred previously ungrouped ICD-9-CM codes to ADGs based on diagnostic patterns found within the Medicare population. (The approach used and actual changes made to the ADG/ACG system as a result of this project are described in ***Chapter III***.)

2. Payment Amounts for Capitated Systems

The second major input of the risk adjuster models developed for this project was the inpatient oriented risk adjuster, Payment Amount for Capitated Systems (PACS). The PACS risk adjuster technology also was developed by physicians, economists, and health service researchers at Johns Hopkins. (See ***Appendix I-3*** for the original PACS components). The original PACS development effort was sponsored by HCFA as a potential prospective methodology for use in paying Medicare risk-contract HMOs.

The original PACS measure included three demographic characteristics (age, sex, and disability status) and three claims-based measures:

- The Major Diagnostic Category (MDC) associated with each hospitalization of a patient in the base year;
- The chronicity of each primary clinical diagnosis that resulted in each patient's hospitalization; and
- The beneficiary's level of overall ambulatory resource use in the base year as measured by whether the Part B Medicare deductible was met.

MDCs were chosen from a range of systems available for classifying inpatient utilization. MDCs are aggregations of Diagnosis Related Groups (DRGs). DRGs are used in the U.S. and abroad as the basis for per case hospital payment. MDCs group patients into one of currently 27 broad, organ-system categories based on the patient's DRG, and ultimately the primary hospital discharge ICD-9-CM diagnosis. For example, all diseases of the nervous system are grouped into MDC 1; all circulatory system diseases are grouped into MDC 5. In the development of the original PACS, certain MDCs were combined to provide a sufficient patient sample size within each MDC included in the model.

MDCs contain diagnoses of both chronic diseases and episodic, self-limited diseases. To capture this dimension, PACS included a three-level chronicity-status variable. The chronicity variable was defined by assigning the patient's principal hospitalization diagnosis into one of the three disease levels: 1) acute; 2) acute with sequelae; or 3) chronic. If a person had multiple hospital admissions, then PACS used the highest chronicity level recorded. The PACS disease chronicity classification system was developed by asking 169 physicians in 31 different specialties to rate each common ICD-9-CM code in their specialty along the three dimensions. In total, the physicians rated over 9,500 diagnosis codes. When the physicians disagreed on the rating of a particular code, a statistical algorithm was used to assign a single level of chronicity to that diagnosis code (Anderson et al. 1989).

The PACS model incorporated ambulatory utilization by determining whether a beneficiary met the Medicare Part B deductible in the preceding year. This helped segregate non users and very low users of ambulatory services (who never reached the modest deductible) from medium and high users. At the time PACS were developed, Part B claims data did not reliably have ICD-9-CM diagnostic information linked to each claim (as it is today), thus clinical ambulatory information was not incorporated into the model.

Given the conceptual overlap of the PACS chronicity assignment and the basic ADG assignment system (which includes a comprehensive assessment of disease patterns), one development task performed during the current project was to compare and assess the PACS chronicity assignments to the disease status designations imbedded within ADG assignments. The project's clinicians evaluated differences between the PACS and ADG chronicity assignments of specific ICD-9-CM codes, and ultimately reassigned some ICD-9-CM codes to different ADGs.

The PACS model developed and investigated using a physician discretion variable. Information on the level of physician discretion for each five digit ICD-9-CM code was collected from the participating 169 physicians via survey. Two components of physician discretion were investigated on the questionnaires. The first component, likelihood of admission, asked the physicians to give their "opinion about the likelihood that an average physician, confronted with a patient with each of the following diagnoses, would admit the patient to the hospital." Physicians rated each ICD-9-CM code at the five digit level on a scale of low, medium, or high likelihood. The second component, discretion in decision to admit, asked the physician to rate the "amount of

variation among physicians in the decision to admit a given patient with a given level of symptoms". Physicians then rated each 5 digit ICD-9-CM code on a scale of low, medium or high discretion (Anderson et al. 89).

The discretion variable was ultimately excluded from the PACS effort because it was not considered to be a sufficiently reliable measure for all patient cases within an ICD-9-CM code. For example, even though tonsillectomies and its diagnoses are often very discretionary, the physician panel could not conclude that all tonsillectomies are discretionary. Our current project revisited the issue of physician discretion, and ultimately excluded, as did PACS, this variable from our risk adjuster models.

The original PACS system contributed an approach for incorporating demographic information and the MDC orientation to capture previous inpatient experiences, as well as the multiple regression analytic approach we followed. The risk adjuster components we considered and the analyses we performed are presented in ***Section B of Chapter II.***

CHAPTER II
METHODS

This chapter describes the methods used to develop the risk adjuster models designed during this project. We begin with a description of our data file development activities, including how we constructed the medical expenditure and diagnostic variables used by our models. The chapter then concludes with a discussion of our model development activities.

A. DATA FILE DEVELOPMENT

Analytical file development for this project was complex. The following section presents a brief discussion of the file development process.

This project used HCFA's recently developed Standard Analytic Files (SAFs), which were derived from the National Claims History File (NCHF). The SAFs contain 100 percent of the institutional bills and a five percent sample of physician/supplier claims. (Ambulatory ICD-9-CM codes became a mandatory part of the "HCFA 1500" claims form in 1989, and were uniformly captured and maintained by HCFA as of 1990.) An assessment of the validity of this data is in ***Appendix II-1.***

The development of the analytic database used for this study involved a high degree of collaboration between the Johns Hopkins/Lewin team and the Diagnosis Cost Group (DCG) risk adjuster development team at Boston University and Center for Health Economics Research. The BU/CHER team developed a revised version of the Diagnostic Cost Group (DCG) system using the same database as the JHU/Lewin team. The BU/CHER and JHU/Lewin research teams constructed a common data set that would allow for a comparison of the results of the two projects. The teams worked with HCFA-ORD staff to agree on common methods for selecting raw data, making data adjustments, and developing master analytic files.

The database development tasks were divided between the two teams, and the resulting claims files were shared. The final database steps and creation of person-level summary records were performed individually by each team. ***Appendix II-2*** documents the technical steps taken to construct the data base. The broader steps are presented here.

1. Base Population

The project data were drawn from five SAFs for the years 1991 and 1992. We used the five percent national file of physician/supplier claims data, and the 100 percent files of inpatient, outpatient, home health and hospice claims. Demographic information on the five percent of Medicare beneficiaries for whom data were retained in the physician supplier databases were obtained from HCFA's HISKEW file. In total, we obtained enrollment and 100 percent of the claims data from a five percent national sample of about 1.5 million beneficiaries.

2. Population Exclusions

Several groups of the Medicare population were excluded from analysis by both teams. Populations were excluded if, due to special circumstances, the majority of their health resource consumption was not documented by the claims data in our possession. The following population groups were excluded from our analysis:

- Persons with no eligibility record matching the claims data;
- Railroad Board retirees;

- Indian Health Service hospitals patients;
- Those not eligible for Medicare Part-B;
- HMO enrollees;
- Residents outside the 50 United States and the District of Columbia;
- Those who were not 65 or older during the first month of the 24 month study period (thus we excluded: under 65 disabled beneficiaries; under 65 end-stage renal disease beneficiaries; and beneficiaries who turned 65 some time during the study period); and
- Beneficiaries who died in year one of the project's two year study period (we retained those who died during the second study year).

After these restrictions were applied, the study population equaled approximately 1.24 million beneficiaries.

3. Split-Half Method

A standard technique used by risk adjustment researchers is to develop a risk adjuster on a "development" database and then test it on a "validation" database. Often, a "split-half" approach is used where data from a single population is large enough, in terms of statistical sampling, to split into two halves. That was the approach that the JHU/Lewin and the BU/CHER teams used. The study population was randomly split in two, so that both risk adjuster teams used the same data split to develop the adjusters, and the same remaining data split to evaluate the adjusters. After the population exclusions described above, each data split included approximately 620,000 individuals. The JHU team created the data file (with the help of the BU/CHER team) and developed the risk adjuster models presented here. The main role of the Lewin team was to act as a semi-independent evaluator of the risk adjusters developed by JHU. Some characteristics of the random split halves of the study population are summarized on ***Table II-1*** below.

TABLE II-1:
COMPARISON OF DEVELOPMENT AND EVALUATION DATA

Characteristic	Development	Evaluation
1992 Annualized payments[a]	$4,266	$4,246
1992 Nonannualized payments	$3,214	$3,207
Percent patients hospitalized	19.9%	19.8%
Percent patients with Hosdom diagnoses[b]	16.4%	16.4%
Percent patients Medicaid eligible	9.9%	9.8%
Percent female	59.8%	59.8%

[a] Annualized payments are the payments after adjustments are made for those with partial-year experience (see below).
[b] Hosdom = "Hospital dominant" diagnoses marker (see ***Chapter III***).

4. Dependent Variable

a. Dependent Variable Construction

The dependent, or outcome, variable used to develop the risk adjusters is Medicare's annual expenditures for each beneficiary. Several options existed in determining the exact

definition of "annual expenditures". The two project teams worked with HCFA-ORD in determining a common, best method for constructing the dependent variables. These methods are summarized below:

- Physician/supplier expenditures were estimated using the resource-based relative value scale (RBRVS) amounts adjusted to reflect local area input prices by the Geographic Practice Cost Index (GPCI) weights for services where RBRVS units have been assigned. Allowed charge amounts were used for services where RBRVS units have not been assigned. (A detailed description of the physician expenditure construction process JHU employed to develop physician expenditures is presented in ***Appendix II-3***);
- Outpatient facility expenditures were estimated using claim payment amounts;
- Inpatient facility expenditures were estimated using DRG payments, plus capital and outlier payments;
- Skilled nursing facility and TEFRA institution (those facilities exempt from prospective DRG payments) expenditures were estimated using claim payment amounts;
- Hospice and home health service expenditures were estimated using claim payment amounts;
- Patient co-payments and deductibles were excluded from the dependent variable; and
- Claims where Medicare was the secondary payor were excluded.

b. Adjustment To Year-Two Decedents

The main goal of this project was to develop a system where diagnostic and demographic information from "year-one" could be used to predict expenditures in "year-two." Persons who were not continuously enrolled in year one because they aged into Medicare or died during that year were excluded from the study. Only continuous enrollees were used so as to have complete base year encounter data. However, to realistically replicate populations in capitated plans, we included beneficiaries who died in year two of the data (about 4.8 percent of the annual Medicare population dies each year).

Both teams used a weighting method developed by BU/CHER in its earlier DCG development to translate the partial year expenses of year two decedents into annual predicted expenditures. ***Appendix II-4*** illustrates the effect of this adjustment of expenditures. The adjustment requires two steps. First, the 1992 expenditure data for these individuals were expressed in annualized dollars. For example, if a person lived for the first six months of 1992 and $25,000 were expended on their behalf during this period, we set the "annualized" expenditures to $50,000. Next, when calculating means and regression coefficients, the observations were weighted back to represent the fraction of year two in which the decedents were living. For example, a weighting factor of .5 would be used for this person. The regression models accordingly would give this person one half the weight of a person who was alive during the entire period.

The formula used to annualize the expenditure data was:

$$Annualized\ Payments = \frac{Estimated\ Payments}{Number\ of\ Months\ Alive\ in\ Year\ Two} * 12$$

When calculating means and regression coefficients, the data observations were weighted back using the formula:

$$\frac{Number\ of\ Months\ Alive\ in\ Year\ Two}{12}$$

c. Outlier truncation

The nature of health care resource consumption is highly skewed, in that a small percentage of persons are responsible for a large proportion of health care expenditures. This results in the presence of outliers in any payment formula, which has a dramatic effect on any risk adjusted expected rate.

The main risk adjuster models presented here were developed without any truncation or transformation. This decision was made in conjunction with HCFA oversight staff to best mimic the Medicare policy contexts and to allow for comparison with past efforts. Unless otherwise noted, the models and reports discussed in this report use models developed with non-truncated data.

In addition, two additional sets of models were developed after truncating annual expenses at $50,000 and at $100,000. That is, individuals with expenditures in excess of the truncation thresholds (either $50,000 or $100,000) were assumed to have expenditures equal to the truncation threshold. The models developed with these truncated data are presented ***in Chapter III***. One use of models estimated with truncated data is to allow analysis of reinsurance schemes. Although reinsurance is not a part of current Medicare policy, a model developed using truncated data is appropriate for plans that include stop-loss reinsurance coverage.

d. Other Dependent Variables

The goal of this project was to develop a capitation adjuster for total payments. It also is possible that an adjuster might be applied to a partial capitation, or blended FFS/capitation, context in which case payment for only certain types of services would be included in the capitation amount. To explore this application, we created several partial payment variables, calculated descriptive statistics and performed regression analyses to assess the degree of predictive power of our risk adjusters for three of the partial payment variables.

The first subset of expenditures regressed was "total physician expenditures", in which we included expenditures for physician services performed in the ambulatory and inpatient settings. The second subset was "total ambulatory expenditures", in which we included physician services performed in the ambulatory setting and outpatient department fees. The final subset analyzed was "total inpatient expenditures", in which we included expenditures for physician services performed in the inpatient setting and inpatient facility fees. (Adjusted R-square statistic results of these subtotal dependent variables are in ***Appendix II-5***).

B. RISK ADJUSTER MODEL DEVELOPMENT

1. Overview and Philosophy

This section describes the modeling activities underlying the final two JHU models presented in ***Chapter III***. In this section, we first explain the overall ADG/ACG diagnosis grouping system and the MDC inpatient stay grouping system. This includes the updates and improvements made to each during the project, as well as the process used to select specific ADG and MDC variables for the final models.

Second, we discuss the integration of the ambulatory, diagnosis-based ADG variables with the inpatient, DRG-based MDC variables that are the basis of the first of the two final models. In addition, we also discuss an additional marker variable for inpatient-oriented diagnoses (the "Hosdom" variable), which is then integrated with the ADG variables in the second model. Thus, our final two models consisted of an ADG-MDC and an ADG-Hosdom model.

Finally, we present the exact claims diagnosis source for the variables of the two final models. Further details on model development are contained in ***Appendix II-6***, while ***Appendix II-7*** offers several examples of models that preceded the final two models.

2. Socio-Demographic Data

Based on JHU's previous work, deliberations by our team, and empirical assessments on the development half of the data, we selected four demographic variables to incorporate into our final adjustment model: age, gender, Medicaid status, and previous disability status. (Exact definitions of these variables are presented in ***Chapter III***).

3. Ambulatory Diagnoses and ADGs

This project is one of the first development projects to apply ambulatory (Part B) diagnosis codes to the task of risk adjustment of Medicare beneficiaries. A major advantage of using these codes to develop our diagnosis-based adjusters is that during each year, over 85 percent of beneficiaries receive ambulatory services and are diagnosed with one or more conditions in this setting. Thus, ambulatory diagnosis-based adjusters can use this information to categorize almost all beneficiaries. This contrasts to the less than 20 percent of beneficiaries who in any given year receive one or more diagnoses in the inpatient setting. While current patterns of care dictate that the sickest (and most expensive) patients are usually treated in an inpatient facility, the trend towards more intensive care provided on an ambulatory basis indicates the importance of an ambulatory element being included in any risk adjuster system.

Our major approach to utilizing diagnoses was to adopt grouping methodologies applicable to all age ranges. Thus, a major objective of this project was to improve ADGs' predictive ability among elderly enrollees while maintaining their applicability among all age cohorts. As such, our design strategy was two-fold: 1) improve the ACG/ADG system's ability to group diagnoses common among the elderly; and 2) identify a parsimonious sub-set of ADGs that best predicts expenditures for the Medicare elderly.

The first step taken to enhance the application of ADGs to the elderly was to add several hundred new ICD-9-CM codes to ICD/ADG "maps" based on codes encountered among elderly in inpatient and ambulatory settings. The original ADG system used about 6,000 of the 10,000 ICD-9-CM codes and accounted for over 95 percent of all ambulatory encounters in the under-65 population.

We expanded the grouping capability for the elderly population by running the original ADG grouper against ICD-9-CM codes found in the development split of the study data, and then categorizing the most frequent "ungrouped" ICD-9-CM codes into the appropriate ADGs based on our original assignment criteria.

In addition, a number of ICD-to-ADG mappings were changed from earlier versions of the ADG system. This reassignment was based on clinical input and empirical analyses of the degree to which specific diagnoses were associated with different types of resource use. (See Weiner et al. 91 for discussion of ICD-to-ADG grouping criteria.) This recategorization was also guided by comparing the original ICD-to-ADG mapping with the three-level chronicity categorization of diagnoses of the original PACS system.

The original ADG morbidity clustering system has undergone criticism and scrutiny over the last 10 years, and is generally considered clinically robust across populations of all ages. However, after analyzing patterns of care and coding among the elderly, our clinical team and consultants modified some original ADG categories. These included, in particular, the three psychosocial ADGs. These ADGs were recategorized into three other groups that better reflected the related

diagnoses and the underlying philosophy of the ACG/ADG system (likelihood of persistence and overall severity).

The research team also assessed the impact of modifying the terminal group ACG algorithm to make it more applicable to the elderly. For example, we added a new ACG with 15 or more ADGs, in order to take into consideration the higher proportion of sicker persons in the Medicare population. We formed the new ACG by analyzing the original top ACG category (composed of 10 or more ADGs), and subdividing it into ACGs composed of patients with 10-14 ADGs and with 15 or more ADGs. In addition, we split ACGs originally composed of several ADGs (and representing generally older and sicker patients), into ACGs composed of only "major" ADG assignments and ACGs composed of only "minor" ADG assignments. Finally, we also explored alternative ACG/age groupings. Ultimately, because an ADG-oriented multivariate model proved to have significantly higher explanatory power among the elderly cohort than did the fixed-cell ACG groups, an ACG-cell model was not developed as a final adjuster. (***Appendix II-8*** shows the percent of variation in total expenditures explained by the cell-based ACG system.)

Extensive iterative modeling was performed on the ADGs to develop the best parsimonious set of significant, positive and stable ADG variables. We explored various models of different ADGs based on clinical judgment, on assessments of patient frequencies, and on statistical significance of model variables at the p= 0.05 level.[1] For example, it is intuitive that a condition like cerebral thrombosis, found in ADG 9 (likely to recur: progressive) is probably going to be more predictive of future resource use than a condition like an upper respiratory tract infection, found in ADG 2 (time-limited minor: infections). It is also intuitive that all variables should have positive coefficients, since all variables represent health system encounters for particular diagnoses versus no encounter for the particular diagnoses. We performed iterative, regular and step-wise regression among five random sub-populations constructed from the development half of the data to eventually develop a parsimonious and clinically acceptable ADG model. The ICD-9-CM diagnosis codes encountered among the study population and their ADG assignments are listed in ***Appendix II-9.*** Using this "mapping" list, we were able to assign an ADG to over 98 percent of the ambulatory and inpatient records we attempted to group into ADGs.

4. Inpatient Data and MDCs

Several modifications of the original PACS were performed in the development of our integrated risk adjusters. The first step was to omit the binary variable indicating whether the Part B deductible was met. This was done given that our risk adjuster system could incorporate the diagnostic ambulatory data now available using ADGs. The second step was to re-estimate the regression equation using the study's 1991 encounter and 1992 expenditure data. This step also allowed us to verify that all common codes currently in use could be grouped by the system, and to categorize new diagnostic codes as necessary. The third step was to incorporate and analyze MDCs that were created since the PACS project (which was developed using mid-1980s data). Ultimately, one new MDC classification (MDC 26 - transplants) was added to our models.

Based on our statistical analyses and clinical review of these analyses, we combined some MDCs into broader categories. For example, MDC 25 (HIV-AIDS) was added to the original PACS grouping of MDCs 16-17 (blood, immunological, and myeloproliferative diseases). MDC 24 (trauma) is a recent classification we eventually combined with MDC 23 (factors influencing health status). Although the frequency in our data set of MDCs 23 and 24 was relatively small, the combined variable of this relatively very ill patient group remained positive and significant.

[1] As it turned out, the ADG (as well as the MDC) variables included in the two final models were significant at the 0.001 confidence level.

The new MDCs described above, along with some of the original MDC categories, were then included in the ADG-MDC model. In addition, five of the original PACS MDCs were also dropped during the development of the ADG-MDC model. The five excluded MDCs were: MDC 12 (diseases and disorders of the male reproductive system); MDC 13 (diseases and disorders of the female reproductive system); MDC 14 (pregnancy, childbirth and the puerperium); MDC 15 (newborns and other neonates with condition originating in perinatal period); and MDC 0 (other). These MDCs either represented few patients or were unstable and statistically insignificant.

To improve the inpatient component of our models, we explored new inpatient diagnosis aggregations by reconstructing MDCs (particularly MDC 12) that had negative, insignificant and/or unstable coefficients. For these problem MDCs, we identified their component DRGs; substituted the individual DRGs for the overall MDC; added these DRGs to the remaining significant and positive MDCs; and re-estimated the model. We then identified and excluded DRGs that were negative and/or insignificant; reincorporated the modified MDCs into our models; and re-estimated the equations. Lacking significant improvement in these modified MDC models, we ultimately simply excluded entire MDCs that were negative, insignificant at the p = 0.05 level, and unstable across random sub-groups of our data.

We also tested three alternative constructions of each MDC variable. To address the concern of possible incentives for over-treating in the inpatient setting, we tested the use of each MDC as "dummy" variables which indicated the presence of no admissions versus one or more hospital admissions within an MDC. We also tested a categorical approach, where admissions were grouped into no admissions, one or two admissions, or three or more admissions within an MDC. We compared these two categorical models with a model using MDCs as count variables – where the MDC coefficient indicates payment for each individual admission, which is multiplied by the number of admissions per year. Among the three formulations of MDCs, the count formulation explained the greatest variability in expenditures, even though 80 percent of the population were never admitted in year one; 13 percent were admitted once; 4 percent were admitted twice; and 3 percent were admitted more than twice. Ultimately, the original, count MDC approach was incorporated in our models.

Finally, we tested the use of the PACS chronicity designation for hospital diagnosis as: 1) a three level variable (chronic, acute with sequelae, acute); 2) a two level variable (chronic, acute); 3) a dummy variable (chronic); and 4) the variable excluded from the model. We ultimately incorporated PACS chronicity information by modifying several ADG assignments, and then excluded the separate hospital diagnosis chronicity variable.

5. Ambulatory/Inpatient (ADG-MDC) Integration

A considerable amount of iterative modeling was performed on ADGs alone; MDCs alone; ADGs with demographic variables; MDCs with demographic variables; ADG and MDCs combined; and ADGs, MDCs and demographic variables combined. Different combinations of these variables were included in a series of preliminary models that were explored by testing several alternative diagnostic claims data sources for ADG and MDC assignment. The final ADG and MDC diagnostic assignment criteria are described later in this chapter. ***Appendix II-10*** illustrates the effect of alternative claims sources for ADG assignment.

As noted, we performed iterative model building using regular as well as step-wise regression techniques, aided by clinical judgment and tests of face validity. We performed these regressions on five random sub-populations constructed from the JHU development half of the data. Across each sample, we then compared the: 1) significance of each variable at the p= 0.05 level; 2) magnitude of each coefficient; 3) direction of each coefficient (positive or negative, indicating amounts added or subtracted from a patient's total predicted payment); and 4) overall

stability across the five sub-samples. We then combined the five sub-samples into one dataset, repeated the iterative regressions and four-step assessment process, and compared the results to work done on the smaller subsamples.

In addition to determining a final set of ADGs and MDCs for inclusion in our risk adjuster models, we tested alternative constructions of the demographic variables included in the models. First, the effect of constructing age in various manners was explored. We compared age as a: 1) continuous variable, as was used in the original PACS inpatient risk adjuster; 2) five-year increment categorical variable, as is used currently in the AAPCC based payments to Medicare's risk based contracts; and 3) ten-year increment variable. We ultimately retained the PACS formulation of age as a continuous variable – it incorporates the greatest amount of information and is most predictive of expenditures of the three alternatives.

We also tested the inclusion of an age/gender interaction variable. This interaction variable tested the possibility of variation in expenditures occurring above and beyond variation in age and gender separately. If age crossed with gender creates a synergistic, or "multiplicative" effect on the annual expenditures of Medicare enrollees, then the variable would be statistically significant. The age/gender interaction variable was not statistically significant, and had a small coefficient when added to our models with age, gender, MDCs and ADGs.

Finally, we assessed the predictive power of models that included : (1) a dummy variable indicating those eligible for Medicare prior to age 65 due to their receipt of disability insurance (DI) benefits; (2) a dummy variable indicating Medicaid eligibility during one or more months in the base year; and (3) a continuous variable indicating the number of months of Medicaid eligibility during the base year. The disability and Medicaid eligibility dummy variables were retained in our final two models; the continuous variable measuring the number of months of Medicaid eligibility was not. The resulting ADG-MDC model is presented in ***Chapter III***.

6. An Alternative Integrated Model: "Hospital Dominant" Diagnoses

In addition to the ADG-MDC approach for capturing diagnoses, this project developed a second approach for incorporating inpatient diagnostic data. The motivation for this second approach was to avoid including in the risk adjuster model any variable that reflected prior utilization in the form of explicit hospitalization. Thus, the motivation for developing a second model was to minimize incentives to hospitalize a patient (to receive a higher capitation payment) for conditions that could also be treated more efficiently in the ambulatory setting.

We termed this second approach the "hospital dominant" (abbreviated as "Hosdom") model, to reflect conditions treated predominately, but not always or necessarily, in the inpatient hospital setting. This second JHU model excluded MDCs, and incorporated inpatient as well as ambulatory diagnosis data into the ADG assignment process and into the binary hospital dominant (Hosdom) variable.

The hospital dominant marker was developed through a several step empirical analysis of the 1.24 million beneficiaries in the combined development and test files. (This is the only development process performed on both the "JHU development" and "Lewin test" datasets.) For every ICD-9-CM diagnosis, we determined the likelihood that a patient received at least some care during the year in an inpatient or ambulatory setting for that condition. Then, we ranked the diagnosis codes based on the proportion of patients that were hospitalized during the year for that diagnosis. Based on this list, we eventually defined the "hospital dominant" conditions to include 843 diagnoses, for which at least 50 percent of patients were hospitalized for that condition once or more during the year. Finally, we performed sensitivity analyses to determine the impact of alternative definitions above the 50 percent cut-off. The percentage of patients hospitalized for the

majority of Hosdom marker diagnoses was much higher than the minimum 50 percent level. However, we used the 50 percent demarcation so as to be more, rather than less, inclusive. A list of all ICD-9-CM codes included on our hospital dominant marker is found in ***Appendix II-11***.

Two risk adjuster variables from the list of ICD-9-CM "Hosdom" codes were considered. The first variable was constructed as a "dummy" variable to reflect the presence of none, versus one or more, "hospital dominant" diagnosis codes in a patient's claims records. With this formulation, a condition is considered serious enough to be usually associated with a hospitalization, even if the patient actually is treated in the ambulatory setting. The second variable created was a count of the number of Hosdom diagnoses found on a patient's claims data. Based on an evaluation of the explanatory power of these two variables, and on the goal of minimizing coding manipulations, we ultimately incorporated the "dummy" Hosdom variable of diagnoses into our second model.

7. Claims Sources For MDC and ADG Assignments

a. MDC Assignment Criteria

The MDC assignment process was straightforward. MDCs (used only in the "ADG-MDC" risk adjuster) were assigned from their component DRGs recorded on inpatient claims. DRGs are assigned to all hospital admissions based on ICD-9-CM diagnosis and procedure codes.

b. ADG Assignment Criteria

The claims source with the diagnostic information used for ADG assignment was more complex and differed for the ADG-MDC and ADG-Hosdom models. The claim's "place of service" field was used to differentiate the two models' ADGs. The "place of service" associated with each health service was used to define each claim's diagnosis codes as "inpatient" or "ambulatory." Both inpatient and ambulatory codes were used to assign ADGs in the ADG-Hosdom model, while only ambulatory codes were used to assign ADGs in the ADG-MDC model.

Inpatient diagnosis codes were derived from several sources. The first source was diagnosis codes noted on the hospital inpatient facility claims (up to 10 codes). In addition, inpatient codes included line-item (up to 10 codes) and header diagnoses (up to four codes) from physician/supplier claims where "inpatient hospital" was the place of service.

In contrast, ambulatory diagnosis codes were derived from a different set of sources. These included diagnoses from hospital outpatient facility claims (up to 10 codes), as well as line-item (up to 10 codes) and header (up to four codes) diagnoses on physician/supplier claims from the following eight places of service:

- office;
- home;
- outpatient hospital department;
- hospital emergency room;
- ambulatory surgical center;
- state and local clinic;
- outpatient rehabilitation clinic; and
- intermediate care facility for the mentally retarded.

c. "Face-to-Face" Procedure Diagnoses

In both the ADG-MDC and ADG-Hosdom models, the diagnostic information used in assigning ADGs from physician claims is recorded only when associated with a service provided by licensed clinicians during "face-to-face" encounters with patients. The rationale for this limitation is to exclude diagnoses that may not have been assigned by a clinician and may be less accurate. For example, diagnosis codes from ancillary services, such as laboratory or radiology services, are not included. These services are more likely to be performed to "rule-out" diagnoses than are face-to-face services. Thus, diagnoses recorded on claims when a simple diagnostic procedure, such as a specific blood test or an electrocardiogram, is the sole purpose of the encounter are not used for ADG assignment.

"Face-to-face" diagnoses are derived using the procedure codes (HCPCS) on the claim. Face-to-face encounters are defined as visits involving an evaluation and/or management service or procedure performed by a physician (MD or DO) or a limited license professional (nurse practitioner, physician's assistant, dentist, podiatrist, social worker, chiropractor, or psychologist). These contacts most frequently take place during a patient visit to a providers' office, or to a hospital outpatient department or emergency room. The range of HCPCS codes used to identify and limit diagnoses to those during face-to-face encounters with providers are:

- 00100 - 01999: anesthesia;
- 10160 - 69979: surgery (excludes maternal care);
- 77261 - 77799: therapeutic radiology;
- 78000 - 79999: nuclear medicine;
- 90701 - 99199: medicine (includes 1991 evaluation and management codes); and
- 99000 - 99499: 1992 evaluation and management.

CHAPTER III
RESULTS: ADG-MDC AND ADG-HOSDOM MODELS

This chapter presents and explains the two risk adjuster models developed by JHU during this project. The first section of the chapter explains the multivariate regression approach to determining capitated payments. The second section discusses and interprets each component, or variable, of the two JHU models. The third section details how each model is used to determine per person capitated payment amounts. Several hypothetical enrollees are used to illustrate this process. The fourth section presents and compares the effects on the percent of variation in expenditures explained by the JHU models when a stop-loss reinsurance mechanism is added to the risk adjustment system. The final section summarizes the data elements necessary for determining capitation rates using the JHU models.

A. EXPLANATION OF MULTIVARIATE REGRESSION RISK ADJUSTER MODELS

Most risk assessment/adjuster models are cell-based. Cell-based models divide a population into a series of mutually exclusive rate cells. Each individual in that insured population is then assigned to one of these cells. Ideally, the expected medical expenditures of individuals within each of these cells should be similar. These rate cells comprise a risk assessment system. Risk adjustment occurs when each individual in a given cell is assigned a value equal to the average expenditures of members of that cell in a payment year.

In contrast, the JHU risk adjustment models are based upon the use of multivariate regression, where demographic and diagnostic information from year one (1991) is used to predict medical expenditures in year two (1992).[1] Multivariate regression is a technique that identifies the independent contribution of each year one risk assessor measure (the demographic and diagnostic variables) to year two medical expenditures. The size of independent contributions of each risk assessment measure is the corresponding estimated coefficient from the regression output.

These estimated coefficients serve as weights for the effects of each risk assessment measure. Risk adjustment occurs when a unique risk score is calculated for each individual. This risk score is the sum of the risk assessment weights for that individual. These risk scores would need to be inflated to a given payment year before risk adjusted payments could be made. The payment calculations for both types of risk adjustment systems are relatively straight forward, however.

Risk adjustment models based on multivariate regression estimates are thus significantly different from rate cell based risk adjustment models. Rate cell approaches use only a few characteristics to specify each rate cell, and then use the average expenditures for each rate cell to predict medical expenditures for each member of an insured population. In

[1] The results we reported here were for ***prospective*** models -- i.e., year one demographic and diagnostic data were used to predict year two medical expenditures. In addition, ***retrospective*** versions of each model were also estimated -- i.e., demographic and diagnostic data in the current year were used to predict medical expenditures in that year (1991). The adjusted R-square statistics for the retrospective models are presented in ***Appendix III-1***.

contrast, regression-based risk adjustment models use multiple characteristics to predict a unique level of medical expenditures for each member of an insured population.

B. INTERPRETATION OF VARIABLES IN JHU MODELS

1. Dependent Variable

The "dependent variable" – the variable predicted by the ADG-MDC and ADG-Hosdom models – is an estimate of 1992 total Medicare payments for each individual. As described in ***Chapter III*** and ***Appendix II-3***, the dependent variable represents DRG payments for inpatient claims; payments for other institutional claims; and RBRVS payment estimates of provider claims. The variable was adjusted to account for individuals enrolled for only part of year two (those that died in 1992) and omitted patient copayments and deductibles.

2. Independent Variables

Each JHU model then used several risk assessors as independent variables in its underlying regression equation. The coefficients from the regression equation output for each variable indicated the risk adjustment score associated with that variable in year two (1992) -- i.e., the expected increase in year two medical expenditures associated with that risk assessor. These risk adjuster annual scores are presented in ***Table III-1*** and ***III-2***. The risk assessors included in each regression equation, or model, are described below.

Both JHU risk adjuster models included four demographic variables. The first was the gender variable, "male." The weight assigned to this variable by each risk adjuster model indicated how much greater the risk score was for males relative to females. The second was an age variable, "years over 65." This was a count variable indicating the number of years each enrollee is over the age of 65. The weight assigned to this variable by each risk adjuster model indicated the amount an individual's risk adjuster score is increased for each year above age 65. For instance, the "years over 65" weight for someone who was age 85 would be multiplied by 20 (i.e., 85 - 65 = 20). The third was the "ever disabled" variable, a binary variable indicating whether a person ever was eligible for Medicare benefits as a result of receiving Social Security Disability Insurance. Individuals who were ever disabled tended to have higher medical expenditures than persons who had never been disabled. The final demographic variable was the "Medicaid" dummy variable. This binary variable indicated whether an individual was eligible for Medicaid benefits in at least one month during year one (1991).

a. The ADG-MDC Model

Both JHU risk adjuster models include the demographic risk assessor variables as described above and an intercept value reflecting a "baseline" individual's expenditures for 1992. The ADG-MDC model (***Table III-1***) predicts that the "base-line" individual's 1992 Medicare expenditures is $608. Based on the demographic variables of our models, a baseline individual is consider a 65 year-old female with no prior disability benefits, no year one Medicaid benefits, and no year-one flags of the models' diagnosis-based variables.

An individual's risk adjuster scores are made relative to the base line case. For example, the ADG-MDC model increases the risk adjuster score by $604 for men; by $67 for each year over age 65; by $1,119 for individuals who were Medicare eligible in the past due to prior disability benefits; and by $761 for those who were Medicaid eligible in year one.

Table III-1
ADG-MDC RISK ADJUSTER MODEL

VARIABLE	PERCENTAGE a/	WEIGHT	STD. ERROR
BASE EXPECTED PAYMENT	NA	$608	$28
MALE	39.2%	604	26
NUMBER OF YEARS OVER AGE 65	10.1 b/	67	2
EVER RECEIVED DISABILITY	6.3%	1,119	51
MEDICAID ELIGIBLE	9.6%	761	43
Visit ADGs (VADGs)			
TIME LIMITED, MAJOR (3)	15.7%	542	36
TIME LIMITED, MAJOR, PRIMARY INFECTIONS (4)	8.4%	734	64
ASTHMA (6)	2.5%	818	123
LIKELY TO RECUR, DISCRETE (7)	23.6%	225	65
LIKELY TO RECUR, PROGRESSIVE (9)	6.7%	965	134
CHRONIC MEDICAL, UNSTABLE (11)	42.0%	1,345	126
CHRONIC SPECIALTY, UNSTABLE, ORTHOPEDIC (16)	2.7%	650	107
INJURIES/ADVERSE EFFECTS, MAJOR (22)	10.0%	525	177
PSYCHIATRIC, TIME LIMITED, MINOR (23)	1.2%	698	110
PSYCHIATRIC, PERSISTENT OR RECURRENT, UNSTABLE (25)	2.8%	804	245
SIGNS/SYMPTOMS, UNCERTAIN (27)	20.2%	460	163
SIGNS/SYMPTOMS, MAJOR (28)	30.7%	551	97
MALIGNANCY (32)	11.1%	1,347	206
Major Diagnostic Categories (MDCs) c/			
NERVOUS SYSTEM (1)	1.9%	$1,533	36
EARS, NOSE, THROAT, RESPIRATORY SYSTEMS (3 or 4)	2.9%	3,237	46
CIRCULATORY SYSTEM (5)	5.5%	1,897	79
DIGESTIVE SYSTEM (6)	2.7%	1,759	30
HEPATOBILLIARY SYSTEM, PANCREAS (7)	0.8%	1,030	53
MUSCULOSKELETAL, CONNECTIVE TISSUE (8)	2.7%	1,117	27
SKIN, SUBCUTANEOUS TISSUE AND BREAST (9)	0.7%	1,762	77
ENDOCRINE, NUTRITIONAL, METABOLIC SYSTEMS (10)	0.8%	2,938	43
KIDNEY, URINARY TRACT (11)	1.0%	2,526	116
INFECTIOUS, PARASITIC DISEASES (18)	0.5%	3,061	79
MENTAL DISEASE, ALCOHOL, DRUG ABUSE (19 or 20)	0.7%	1,957	32
INJURIES, POISONINGS, BURNS (21)	0.2%	1,882	29
HEALTH STATUS FACTORS, TRAUMA (23 0r 24)	0.5%	1,481	40
BLOOD, IMMUNOLOGICAL, MYELOPROLIFERATIVE DISEASES, HIV, AIDS (16, 17, or 25)	0.4%	3,875	79
TRANSPLANTS (26)	0.4%	3,944	60

a/ Percentage of persons within indicated category.

b/ This statistic represents the mean number of years above age 65.

c/ The percentages for the MDC variables indicate the percentage of the population from the development sample who had at least one admission in that MDC in year one (1991). Of those hospitalized within an MDC in year one, however, 92 percent had only one admission in that MDC.

For its diagnosis-based variables, the ADG-MDC model uses 15 MDC groupings. This risk adjuster model's MDCs are count variables that indicate an individual's risk score for each inpatient admission that individual had in year one in a particular MDC. For example, the ADG-MDC model indicates that an individual's risk score in 1992 would increase by $1,533 for each admission that person had in MDC 1 - nervous system in 1991.

Both the ADG-MDC and ADG-Hosdom risk adjuster models incorporate 13 selected ADG groupings. However, the claims source of diagnostic information in these 13 ADGs differs in the two JHU models. The ADG-MDC risk adjuster uses 13 ambulatory "Visit ADGs" (VADGs). VADGs refer to ADGs that are assigned from diagnoses (either primary or secondary) noted by providers during "face-to-face" encounters[2] in the ambulatory setting[3].

Unlike the MDC "count" variables, each VADG is a binary variable that can be triggered only once during the base year, regardless of the number of diagnoses an individual may have in each VADG. For example, VADG 3 (which clusters diagnoses that are time limited, but major) is associated with an increase in year-two individual capitation payments of $542, regardless of the number of similar diagnoses or visits that a patient had during year-one.

b. The ADG-Hosdom Model

The ADG-Hosdom risk adjuster model (***Table III-2***) uses 13 "All ADGs" (ALADGs.) ALADGs use ambulatory and inpatient diagnoses. Specifically, they refer to ADGs that are assigned from all primary and secondary diagnoses noted on inpatient and outpatient facility claims, as well as those noted by clinicians during face-to-face encounters in both the ambulatory and inpatient settings.

The ADG-Hosdom risk adjuster model also incorporates the new "Hospital Dominant" marker. The hosdom marker is a binary variable indicating the presence within an individual's claims records of one or more of 843 ICD-9-CM codes that are serious enough to usually be treated on an inpatient basis. If the marker is triggered, then a payment weight of $1,749 is applied (only once) when summing scores to calculate an individual's annual capitation payment amount. The hosdom amount is in addition to the weight of the ADG in which the hosdom diagnosis may fall.

---

[2] As described in ***Chapter II***, "face-to-face encounters" are defined as visits involving an evaluation and/or management service or a procedure performed by a physician (MD or DO) or a limited license professional (nurse practitioner, physician's assistant, dentist, podiatrist, social worker, chiropractor, or psychologist). A range of HCFA Common Procedure Coding System (HCPCS) procedure codes (which is an expansion of the CPT-4 system) was used to identify and limit diagnoses to those during face-to-face encounters. The procedure code ranges are: (1) 00100 - 01999 for anesthesia; (2) 10160 - 69979 for surgery (excluding maternal care); (3) 77261 - 77799 for therapeutic radiology; (4) 78000 -79999 for nuclear medicine; (5) 90701 - 99199 for medicine (includes 1991 evaluation and management codes); and (6) 99000 - 99499 for 1992 evaluation and management codes.

[3] As described in ***Chapter II***, diagnosis codes designated as "ambulatory visit" codes derive from all available line-items and header diagnoses from hospital outpatient facility claims, and all available line-item and header diagnoses (four maximum of each) from physician/supplier claims -- that are associated with one or more of eight ambulatory-oriented sites of service. These sites of service (on the HCFA physician/supplier file) are: (1) office; (2) home; (3) outpatient hospital department; (4) hospital emergency room; (5) ambulatory surgical center; (6) state and local clinic; (7) outpatient rehabilitation clinic; and (8) intermediate care facility for the mentally retarded.

Table III-2
ADG-HOSDOM RISK ADJUSTER MODEL

VARIABLE	PERCENTAGEa/	WEIGHT	STD. ERROR
BASE EXPECTED PAYMENT	NA	$434	$28
MALE	39.2%	613	26
NUMBER OF YEARS OVER AGE 65	10.1b/	64	2
EVER RECEIVED DISABILITY	6.3%	1,176	52
MEDICAID ELIGIBLE	9.6%	802	43
All ADGs (ALADGs)			
TIME LIMITED, MAJOR (3)	18.6%	663	35
TIME LIMITED, MAJOR, PRIMARY INFECTIONS (4)	9.9%	1,503	44
ASTHMA (6)	2.7%	1,216	76
LIKELY TO RECUR, DISCRETE (7)	25.3%	365	30
LIKELY TO RECUR, PROGRESSIVE (9)	8.3%	1,696	49
CHRONIC MEDICAL, UNSTABLE (11)	44.2%	1,415	27
CHRONIC SPECIALTY, UNSTABLE, ORTHOPEDIC (16)	2.9%	593	74
INJURIES/ADVERSE EFFECTS, MAJOR (22)	11.9%	462	40
PSYCHIATRIC, TIME LIMITED, MINOR (23)	1.4%	1,222	107
PSYCHIATRIC, PERSISTENT OR RECURRENT, UNSTABLE (25)	3.7%	1,088	69
SIGNS/SYMPTOMS, UNCERTAIN (27)	21.5%	568	32
SIGNS/SYMPTOMS, MAJOR (28)	33.5%	753	30
MALIGNANCY (32)	11.5%	1,429	40
Hosdom Marker			
Probable Hospitalization Diagnosis	16.4%	1,749	43

TABLE III-3:
DEMOGRAPHIC ("AAPCC") MODEL

VARIABLE	PERCENTAGEa/	WEIGHT	STD. ERROR
BASE EXPECTED PAYMENT	NA	$1,893	$26
MALE	39.2%	$733	$26
NUMBER OF YEARS OVER AGE 65	10.1b/	$108	$2
EVER RECEIVED DISABILITY	6.3%	$1,895	$53
MEDICAID ELIGIBLE	9.6%	$1,316	$45

a/ Percentage of persons within indicated category.
b/ This statistic represents the mean number of years above age 65.

c. The "AAPCC"

For comparison purposes during this project, we tested a regression-based risk adjuster model approximating the components of the AAPCC method of paying Medicare risk contracts. The AAPCC makes HMO-specific adjustments for age, gender, Medicaid status, and nursing home residence status. The nursing home residence status data were not available in the study data. Instead, our comparison model includes the four socio-demographic components as constructed for the project's two risk adjustment models: gender, age, Medicaid eligibility status (e.g., welfare

status), and prior disability status. Although the AAPCC includes age and gender as a series of five year age group dummy variables for each gender, we included gender as a binary variable and age as a continuous variable. ***Table III-3*** presents this "AAPCC" comparison model.

d. Base Expected Payment

One variable included in all three risk adjuster models that has not been discussed is the intercept term, referred to above as reflecting a baseline individual. In all three models, this baseline individual was a 65 year old female who was not eligible for Medicaid benefits in one or more months in year one (1991), nor Medicare eligible in the past due to a disability.

This baseline, however, also differed between the three models. The largest baseline group of the three models was that of the "AAPCC" model, which included all individuals described in the previous paragraph. The baseline groups for the ADG-MDC and ADG-Hosdom models, however, are smaller. In the ADG-MDC model, the baseline person also did not have any of the 13 visit ADGs nor any of the 15 MDC variables in year one (1991). Similarly, the baseline individual in the ADG-Hosdom model did not have any of the 13 "all ADGs" nor any hosdom diagnosis in year one. It was these further diagnostic restrictions that made the baseline groups for the two JHU models smaller than that of the "AAPCC" model. It is also worth noting that the baseline groups for the ADG-MDC and ADG-Hosdom models were not the same, because the diagnostic restrictions for each were different.

C. ILLUSTRATION OF CAPITATION RATE DETERMINATION USING JHU MODELS

JHU's risk adjuster models assign each Medicare enrollee a unique risk score (measured in year two (1992) dollars) depending on his or her demographic and diagnostic data. While regression-based risk adjuster models are new and rely on a more complex statistical method than cell-based risk adjuster models, using a regression-based risk adjuster model to calculate individual risk adjuster scores is a matter of simple addition. ***Table III-4*** illustrates the arithmetic necessary to determine annual capitation rates for seven hypothetical health plan enrollees. The table presents seven enrollees with varying morbidity levels and health system encounters; calculates the capitation rate of each patient based on his or her socio-demographic and diagnostic assignments; and compares the capitation rates determined from the two JHU models and the comparison ("AAPCC") model.

To determine the capitated rate for individuals for each risk adjuster model, one refers to the models' regression estimates found in ***Tables III-1*** through ***III-3***. For example, "Enrollee 1" in ***Table III-4*** represents an 85 year old male who had no ambulatory encounters or inpatient admissions during year one (1991). For the ADG-MDC model (***Table III-1***), one would add the amount for men ($604), and for someone age 85 (20 years over age 65 * $67 = $1,340) to the intercept ($608), for a total of $2,552 in year two (1992). Based on entries in ***Table III-2,*** the ADG-Hosdom model predicted this individual's year two expenditures to be $2,327. Based on entries in ***Table III-3,*** the "AAPCC" model predicted the expenditures for this hypothetical healthy 85 year old male to be $4,785.

The hypothetical individuals presented in ***Table III-4*** illustrate an important difference between the JHU risk adjusters and the "AAPCC" model. Because the "AAPCC" model included no risk assessors that directly control for health encounters and diagnoses in year one, this model predicted the same health expenditures for all individuals with the same age, gender, prior disability and Medicaid eligibility status regardless of their diagnoses. Thus, the "AAPCC" model predicted, for example, that all 85 year old males who were never disabled and were not eligible for Medicaid

Table III-4:
Determining Capitation Rates for Seven Health Plan Enrollees

	ADG-MDC	ADG-HOSDOM	"AAPCC"
Enrollee 1: No Health System Encounters			
Male	$604	$613	$732
85 years (20 years * Payment Weight)	$1,340	$1,280	$2,160
Base Cost (model intercept)	$608	$434	$1,893
Capitation Rate	**$2,552**	**$2,327**	**$4,785**
Enrollee 2: No MDCs, 1 ADG			
Male	$604	$613	$732
85 Years	$1,340	$1,280	$2,160
Base Costs (model intercept)	$609	$434	$1,893
Doctor visit for depression (ADG 23)	$698	$1,222	
Capitation Rate	**$3,250**	**$3,549**	**$4,785**
Enrollee 3: No MDCs, 3 ADGs			
Male	$604	$613	$732
85 years	$1,340	$1,280	$2,160
Base Cost (model intercept)	$608	$434	$1,893
Depression (ADG 23)	$698	$1,222	
Gastric Ulcer (ADG 7)	$225	$365	
Coronary Atherosclerosis (ADG 11)	$1,345	$1,415	
Capitation Rate	**$4,820**	**$5,329**	**$4,785**
Enrollee 4: No MDCs, 3 ADGs, Hosdom Marker			
Male	$604	$613	$732
85 Years	$1,340	$1,280	$2,160
Base Cost (Model intercept)	$608	$434	$1,893
Depression (ADG 23)	$698	$1,222	
Gastric Ulcer (ADG 7)	$225	$356	
Coronary Atherosclerosis (ADG 11)	$1,345	$1,415	
HOSDOM diagnosis marker	$0	$1,749	
Capitation Rate	**$4,820**	**$7,078**	**$4,785**
Enrollee 5: No MDCs, 6 ADGs			
Male	$604	$613	$732
85 Years	$1,320	$1,280	$2,160
Base Cost (Model intercept)	$608	$434	$1,893
Depression (ADG 23)	$698	$1,222	
Gastric Ulcer (ADG 7)	$225	$365	
Coronary Atherosclerosis (ADG 11)	$1,345	$1,415	
Corneal Edema (ADG 3)	$542	$663	
Diabetes (ADG 9)	$965	$1,696	
Heart Palpitations (ADG 27)	$460	$568	
Capitation Rate	**$6,787**	**$8,256**	**$4,785**

TABLE III-4: (Continued)
DETERMINING CAPITATION RATES FOR SEVEN HEALTH PLAN ENROLLEES

	ADG-MDC	ADG-HOSDOM	"AAPCC"
Enrollee 6: 2 MDCs, 6 ADGs, Hosdom Marker			
Male	$604	$613	$732
85 Years	$1,340	$1,280	$2,160
Base Cost (Model intercept)	$608	$434	$1,893
Depression (ADG 23)	$698	$1,222	
Gastric Ulcer (ADG 7)	$225	$365	
Coronary Atherosclerosis (ADG 11)	$1,345	$1,415	
Corneal Edema (ADG 3)	$542	$663	
Diabetes (ADG 9)	$965	$1,696	
Heart Palpitations (ADG 27)	$460	$568	
2 Circulatory admissions (MDC 5 x 2)	$3,794	$0	
or 1 Hosdom diagnosis marker	$0	$1,749	
Capitation Rate	**$10,581**	**$10,005**	**$4,785**
Enrollee 7: 4 MDCs, 6 ADGs, Hosdom Marker			
Male	$604	$613	$732
85 Years	$1,340	$1,280	$2,160
Base cost (intercept)	$608	$434	$1,893
Depression (ADG 23)	$698	$1,222	
Gastric Ulcer (ADG 7)	$225	$365	
Coronary Atherosclerosis (ADG 11)	$1,345	$1,415	
Corneal Edema (ADG 3)	$542	$663	
Diabetes (ADG 9)	$965	$1,696	
Heart Palpitations (ADG 27)	$460	$568	
2 Circulatory Admissions (MDC 5x2)	$3,794	$0	
or 1 Hosdom diagnosis Marker	$0	$0	
2 Respiratory Admissions (MDC 3x2)	$6,474	$0	
Capitation Rate	**$17,055**	**$10,005**	**$4,785**
GROUP CAPITATION RATE (Sum of all seven patients)	**$49,865**	**$46,549**	**$33,495**

in year would have had medical expenditures of $4,785 in year two (1992) regardless of their health system encounters.

In contrast, both JHU risk adjuster models predicted higher expenditures in year two for individuals with more health encounters and ADG diagnoses in year one. For example, while enrollees one and two are both 85 year old males who were never disabled and were not Medicaid eligible in year one, both the ADG-MDC and ADG-Hosdom models predicted higher year expenditures in year two for enrollee two ($3,250 and $3,459, respectively) than they did for enrollee one ($2,552 and $2,327, respectively). This is because enrollee two had one ADG in year one (one or more ambulatory diagnoses for depression (ADG23)).

There were also important differences between the two JHU models. For instance, the ADG-Hosdom model generally resulted in higher year two capitation rates for enrollees who were not hospitalized in year one (such as enrollees 2-5). This occurred because the ADG-Hosdom model averaged the year two medical expenditures of those who were and were not hospitalized in year one. This happens because the hosdom and ALADG variables include patients that may or may not be hospitalized. In contrast, the ADG-MDC model included the MDC variables that captured some of the differences in the medical expenditures in year two between those who were and were not hospitalized in year one. For enrollees who were hospitalized (such as enrollees 6-7), the ADG-MDC model resulted in higher year two capitation rates.

D. SIMULATION OF STOP-LOSS REINSURANCE MECHANISMS

Most managed health care plans employ a stop-loss reinsurance mechanism to protect against catastrophic health expenditures. To approximate a reinsurance mechanism at the $100,000 and $50,000 levels, the two JHU (and "AAPCC" comparison) models were re-estimated after 1992 expenditure data were truncated at $100,000, and were re-estimated again after data were truncated at $50,000. In other words, in the first re-estimation, individual 1992 expenditures that exceeded $100,000 were assigned a value of $100,000. In the second re-estimation, individual 1992 expenditures that exceeded $50,000 were assigned a value of $50,000. Detailed estimates for those "truncated" models are presented in ***Appendix III-1***. ***Appendix III-2*** presents comparable information for retrospective models where 1991 information is used to predict 1991 expenditures.

As shown below, simulations of reinsurance mechanisms increase the ability of risk adjuster models to explain the variation in medical expenditures on an individual enrollee level. This increased explanation occurs because, especially at the individual level, catastrophic events are often random and difficult to predict with any diagnosis-based risk adjuster model.

Table III-5 summarizes and compares the amount of total variation in individual level enrollee expenditures explained by the JHU and comparison models when 1992 expenditures are not truncated, truncated at $100,000, and truncated at $50,000.

Table III-5:
Percent of Variation Explained
in 1992 Total Individual Expenditures

Model	Not Truncated a/	Truncated at $100,000	Truncated at $50,000
ADG-MDC	6.3%	8.0%	9.0%
ADG-Hosdom	5.5%	7.0%	8.0%
"AAPCC"	1.0%	1.3%	1.6%

a/ The results (adjusted r-square) statistics reported in this table were based on the development sample.

Using untruncated data, the ADG-MDC model explains 6.3 percent of total variation at the enrollee level in annual medical expenditures. The ADG-Hosdom model explains 5.5 percent of total variation in annual medical expenditures. The "AAPCC" comparison model explains only 1.0 percent of total variation in annual medical expenditures. It has been estimated that only 12 to 18 percent of the total variation in individual medical expenditures can be explained (Newhouse 91). According to this upper limit of 12 to 18 percent, the proportion of "explainable" variance in annual

expenditures accounted for by the ADG-MDC model, for example, could be 35 to 52 percent (6.3% / 18% = 35% and 6.3% / 12% = 52%.)

When the data are truncated at $100,000 to approximate the effects of a stop-loss mechanism at that level, a one-third increase occurs across all models in the proportion of total variation in medical expenditures explained. (For example, with the ADG-MDC model, 8% of expenditures explained with truncated data represents a one-third increase over 6.3% of expenditures explained using un-truncated data).

When the data are truncated at the $50,000 level, a 43 to 45 percent increase occurs across the JHU models in the proportion of total variation in medical expenditures explained. (For example, with the ADG-MDC model, 9% of expenditures explained with truncated data represents a 43 percent increase over 6.3% of expenditures explained using un-truncated data).

E. DATA REQUIREMENTS FOR JHU MODELS

The project's risk adjuster models (and accompanying software) are designed to use the data typically retained in machine-readable health insurance claims, encounter data, or enrollment files. For each patient, variable assignment and patient pricing can be accomplished by inputting raw claims data, or by constructing a minimal dataset composed of the following data elements:

- A unique identifier for every member eligible to use health plan services;
- The age, date of death (if applicable), gender, Medicaid eligibility status, and prior disability status of each member;
- All ICD-9-CM diagnosis codes assigned by providers and facilities for all encounters during the time period in question;
- All DRGs assigned during inpatient admissions during the time period in question (for the ADG-MDC model only);
- An expenditure measure, such as allowed charges, from each claim line item (expenditure measures are necessary only if capitation weights are to be developed from the population, rather than from national or HCFA-provided weights);
- Place of service indicators; and
- Procedure (HCPCS) codes.

CHAPTER IV
EVALUATING AND TESTING THE JHU RISK ADJUSTER MODELS

This chapter describes the Lewin-VHI evaluation of the JHU risk adjuster models, in comparison with the "AAPCC" model. Researchers that design or evaluate risk adjuster models use several evaluation criteria (Hornbrook 1991, Epstein and Cumella, 1988, and Anderson et al. 1986), including:

- prediction of costs at an individual level;
- prediction of the costs for groups of individuals, including groups that might be enrolled by plans on a biased basis;
- resistance to manipulation;
- appropriate incentives for efficiency;
- applicability to the entire population of potential participants -- i.e., does the model "work" for all potential Medicare eligibles; and
- administrative feasibility, including the information requirements the system places on providers.

BU/CHER and Lewin-VHI/JHU worked with HCFA-ORD to select statistical measures to apply to population groups to test for predictive accuracy of the respective risk adjuster models at both an individual and group level. First, we discuss the various measures and tests we used to assess the predictive accuracy of the two JHU risk adjuster models. Next, we report the results of our statistical evaluation.

A. MEASURES AND TESTS USED TO EVALUATE THE PREDICTIVE ACCURACY OF RISK ADJUSTER MODELS

The Lewin-VHI evaluation team used two tests to assess the predictive accuracy of the two JHU Medicare risk adjuster models. These tests were applied to numerous population subsamples created from the evaluation half sample of the Medicare data.

1. THE TYPES OF TEST

a. Adjusted R Square Statistic

The adjusted R square statistics, our first test, measures each risk adjuster model's individual level predictive accuracy. As indicated in ***Chapters II and III***, each risk adjuster model predicts the dependent variable medical expenditures in year two (1992) as a function of an individual's demographic characteristics (age, gender, whether the person ever was Medicare eligible due to disability, and whether the person was currently eligible for Medicaid benefits) and medical history (the ADGs, MDCs, and the hospital dominant variable) in year one (1991).

This basic structure is used to formulate regression equations. The R square statistic from these regression equations provides a test of the predictive accuracy of a given risk adjuster model. The R square statistic measures the fraction of the total variance in year two medical expenditures *at an individual level* accounted for by a particular risk adjuster model.

The R square statistics presented in this report are *adjusted* R square statistics and were calculated by reestimating each risk adjustor model for the non-random groups selected from the ***test*** split-half sample. The R squared statistic from a regression equation will increase or remain the same as the number of independent variables included in the regression equation increases. In contrast, the adjusted R square statistic balances increases the model's overall explanatory power against the increase in the number of variables in the model. Strictly speaking, the adjusted R square statistic increases when the t-statistic of any variable added to a model is greater than 1.00.

The adjusted R square statistic varies across risk adjuster models for a given population. These differences indicate the relative ability of each model to account for the variation in year two medical expenditures for individuals within that population. In addition, the adjusted R squared statistic will also vary across different population groups for a given risk adjuster model.

In theory, one uses the adjusted R square statistic to compare the performance of different risk adjuster models across a particular population. The model with the highest adjusted R square statistic for a given population explains the highest fraction of the variance in medical expenditures for that group. The model with the highest adjusted R square statistic for a given population can thus be considered the "best" risk adjuster model for that population using this test. [1]

Risk adjuster models with high adjusted R square statistics are thought to be more resistant to undetected selection effects. Higher adjusted R square statistics reflect that a risk adjuster model does a better job at estimating the expenditures of each individual. This implies that plans will find it less rewarding to "cherry-pick" the best risks relative to a given risk adjuster system and avoid enrolling (or disenrolling) the poorer risks, because payments will more closely track patients' actual medical expenditures.

b. The Predictive Ratio

Risk adjuster models will be used to formulate premium payments or capitation rates to plans based on the risk of all the plan's enrollees. Thus, a superior risk adjuster model should do better at predicting the costs of enrollee groups as well as predicting better at an individual level.

A common "group" test statistic is the predictive ratio. The predictive ratio is defined as:

$$\frac{\sum_{i \in G} Expected\ Expenditures_i}{\sum_{i \in G} Actual\ Expenditures_i}$$

where "i∈G" indicates the sum over all members of group G. This is the ratio of the total expected expenditures for group G to the total actual expenditures for group G. A value of 1.00 indicates that the risk adjuster model exactly predicts the expenditures in year two for group G. Values in excess of 1.00 indicate the risk adjuster model overestimates the true expenditures for group G, and values less than 1.00 indicate the model underestimates the true expenditures for group G. These predictive rations were calculated for random and non-

[1] Strictly speaking, R-square statistics, not adjusted R-square statistics, indicate the fraction of the variance in a regression equation's dependent variable accounted for by the equation's independent variables. We preferred reporting adjusted R-square statistics in this report, however, because of their adjusted R-square statistics account for the effect of adding additional independent variables to a regression model.

random groups selected from the test split-half sample using estimates for each model based on the entire development split-half sample.

2. THE POPULATIONS USED TO TEST A GIVEN MODEL

a. Repeated Random Subsamples

We used two types of population groups to test each risk adjuster model. The first set of groups was repeated random subsamples of the test subsample of different sizes. Specifically, we have tested each model for 100 groups of the following sizes: (1) 500; (2) 1,000; (3) 5,000; (4) 10,000; and (5) 50,000. For these random groups, we calculated predictive ratios for the ADG-MDC and ADG-Hosdom models and for our "AAPCC" model. Our analysis of random groups also includes a simulation of the effects of including individual stop-loss reinsurance with the risk adjustment models.

b. Non-Random Groups

The second set of groups used to test the risk adjuster models was non-random groups. Testing performance on non-random groups provided important additional information on the relative performance of different risk adjuster models. For instance, a given risk adjuster model may be particularly well-suited to predict the expenditures of a given group defined by age and gender but less able to predict the expenditures for the remaining age/gender groups. It may also be the case that a given risk adjuster model consistently over-estimates the medical expenditures of some groups while consistently under-estimating those of another group.

Such information provides a number of useful insights. First, model designers might alter their model if the model does a poor job of predicting the expenditures of some non-random groups. Second, insurers may decide to market more aggressively to individuals in groups for whom a given risk adjuster model consistently overestimates expenditures and to avoid those groups for whom the model consistently underestimates expenditures. A critical goal of risk adjustment is, of course, to reduce incentives for the latter behavior.

The following list of non-random groups was jointly developed by Lewin-VHI/JHU and HCFA ORD for inclusion in the analysis:

- **Age/gender cells** -- where the age (in year one -- 1991) are 65 to 69, 70 to 74, 75 to 79, 80 to 84, and 85+;
- **Race** -- white, black, and other;
- **Hospital admissions** -- one, two, and three or more hospital admissions;
- **Heavy users of ambulatory services** -- these are individuals with no hospital visits but whose use of physician services is high (i.e., their use of physician relative value units (RVUs) is in excess of one standard deviation above the mean number RVUs);
- **Medical conditions** -- enrollees with one or more of the following 17 medical conditions: (1) depression; (2) alcohol and drug abuse; (3) hypertensive heart/renal disease; (4) benign/unspecified hypertension; (5) diabetes with complications; (6) diabetes without complications; (7) heart failure/cardiomyopathy; (8) acute myocardial infarction; (9) other heart disease; (10) chronic obstructive pulmonary disease; (11) colorectal cancer; (12) breast cancer; (13) lung/pancreas cancer; (14) other stroke; (15) intracerebral hemorrhage (16) hip fracture; and (17) arthritis;

- **Expenditure groups** -- quintiles of annualized payment -- the first quintile group has the lowest annual medical expenditures in 1991, while the fifth quintile group had the highest; and
- **Geographic location** -- the nine census division regions.

In the text, we present results for non-random groups defined using year one (1991) individual characteristics. For instance, the expenditure quintile groups were defined using annualized payments in 1991, and the chronic conditions groups were similarly defined using diagnoses recorded in 1991. Alternatively, these groups could be defined using year two (1992) individual characteristics. We have calculated the year two results and present them in ***Appendix IV-2.***

The reason we have not included the results for the groups defined using year two individual characteristics in this chapter is that assessing the performance of prospective risk adjuster models across these groups is not a reasonable test. While year one individual characteristics are known to prospective risk adjustment models and can be used to predict year two expenditures, year two individual characteristics are not.

These non-random groups will provide information on the relative ability of each risk adjuster model to limit HMO selection bias. The performance of each risk adjuster model across these groups will indicate how well a given model might operate for groups differentiated by age, gender, and race, for groups with low or high use of medical services and/or medical expenditures, and for groups in different regions of the county. It is important to note, however, that these groups represent extreme cases -- i.e., what happens if HMOs enrolled only individuals with a certain medical condition or those who were especially low or high users of medical services. In practice, such extremes would be rare or non-existent.

B. RESULTS

In this section, we first describe our evaluation results for the repeated random samples. Next, we proceed with a discussion of our findings for the non-random population groups.

1. RANDOM SAMPLES

a. Risk Adjusters Alone

In ***Table IV-1,*** we present the distribution of the predictive ratios for the "AAPCC," ADG-MDC, and ADG-Hosdom risk adjuster models for each of the five sample sizes (i.e., 500, 1,000, 5,000, 10,000, and 50,000). We include only predictive ratios for the repeated random samples, because individual variation in expenditures (the adjusted R square tests) are not relevant for random samples. We also present these results for groups of size 500, 5,000, and 50,000 graphically in ***Charts One through Three.***

Ideally, the predictive ratios for these random samples should cluster closely around 1.00 for all three models. Due to a law of large numbers one would expect that the clustering around 1.00 would increase as the size of the random samples grows larger.

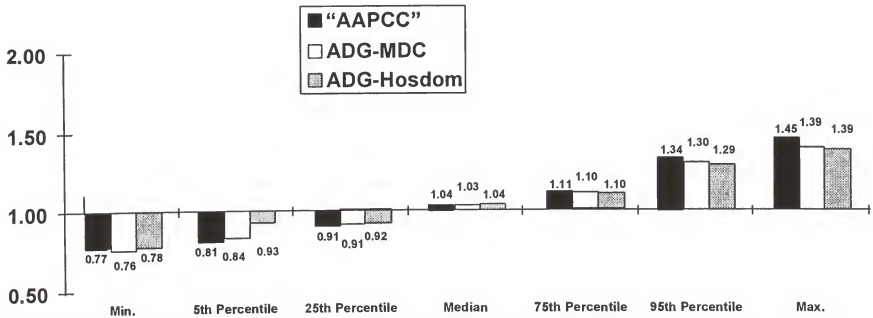
The results for these repeated random samples are mixed. On one hand, the median and mean predictive ratios both cluster near 1.00 for all random group sizes and all three risk adjuster models. The predictive ratios tend to cluster more tightly around 1.00 for the two JHU

TABLE IV-1: DISTRIBUTION OF PREDICTIVE RATIOS FOR REPEATED RANDOM SAMPLES											
Model	Minimum	5th Percentile	10th Percentile	25th** Percentile	Median	75th** Percentile	90th Percentile	95th Percentile	Maximum	Mean	Std. Dev.
100 Groups of 500 Individuals											
"AAPCC"	0.7699	0.8115	0.8529	0.9073	1.0367	1.1140	1.2611	1.3352	1.4528	1.0370	0.1552
ADG-MDC	0.7613	0.8355	0.8441	0.9092	1.0312	1.1047	1.2394	1.2968	1.3924	1.0328	0.1414
ADG-Hosdom	0.7756	0.8304	0.8513	0.9169	1.0370	1.0977	1.2222	1.2879	1.3850	1.0317	0.1390
100 Groups of 1,000 Individuals											
"AAPCC"	0.6865	0.8289	0.8460	0.9133	1.0302	1.1390	1.2377	1.2772	1.3162	1.0341	0.1405
ADG-MDC	0.7493	0.8371	0.8509	0.9298	1.0263	1.1279	1.2046	1.2272	1.2726	1.0294	0.1260
ADG-Hosdom	0.7441	0.8379	0.8625	0.9333	1.0236	1.1217	1.1944	1.2351	1.2918	1.0287	0.1237
100 Groups of 5,000 Individuals											
"AAPCC"	0.8188	0.8590	0.8784	0.9175	0.9776	1.0521	1.1483	1.1698	1.2277	0.9946	0.0982
ADG-MDC	0.8444	0.8602	0.8846	0.9297	0.9839	1.0593	1.1213	1.1477	1.1647	0.9951	0.0850
ADG-Hosdom	0.8564	0.8683	0.8880	0.9391	0.9892	1.0582	1.1154	1.1444	1.1751	0.9970	0.0821
100 Groups of 10,000 Individuals											
"AAPCC"	0.8387	0.8602	0.8774	0.9214	0.9867	1.0700	1.1301	1.1767	1.2319	1.0027	0.0956
ADG-MDC	0.8452	0.8780	0.8864	0.9484	0.9899	1.0611	1.1123	1.1409	1.1733	1.0027	0.0806
ADG-Hosdom	0.8448	0.8866	0.8942	0.9522	0.9980	1.0587	1.1129	1.1411	1.1712	1.0039	0.0774
100 Groups of 50,000 Individuals											
"AAPCC"	0.8866	0.8901	0.8923	0.9040	1.0040	1.0461	1.1086	1.1325	1.1607	0.9994	0.0777
ADG-MDC	0.8999	0.9063	0.9091	0.9344	0.9972	1.0545	1.0877	1.1127	1.1294	0.9990	0.0665
ADG-Hosdom	0.9072	0.9129	0.9144	0.9363	1.0002	1.0410	1.0852	1.1139	1.1360	0.9998	0.0638

* The "mean" predictive ratio is the average of the 100 predictive ratios for each group size.

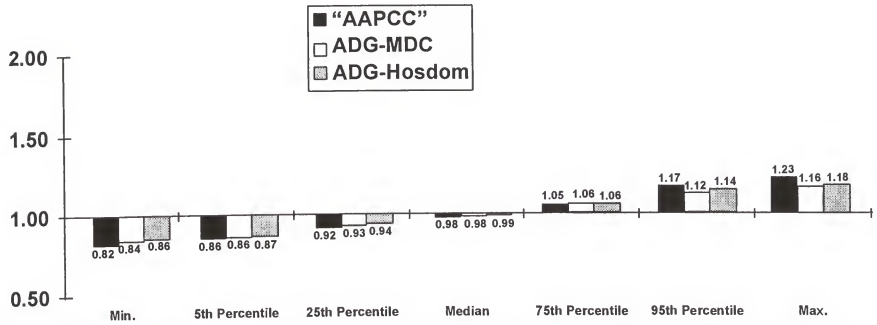
** The difference between the 25th and 75th percentiles is the inter-quartile range.

CHART ONE
DISTRIBUTION OF PREDICTIVE RATIOS FOR RANDOM SAMPLES OF SIZE 500



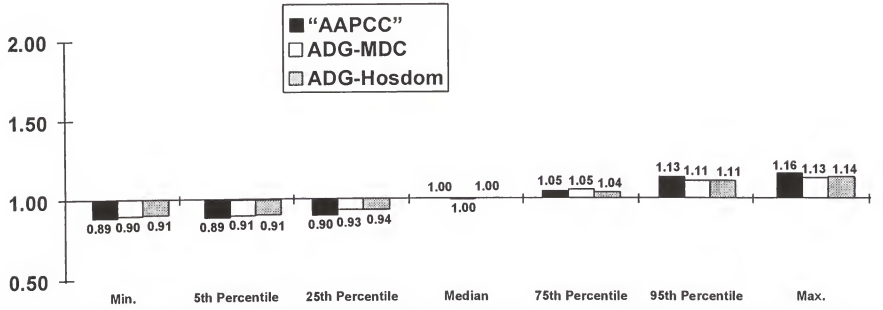
Note: A "perfect" predictive ratio of 1.00 indicates that the risk adjuster model exactly predicted the expenditures of a given group. Predictive ratios of less than 1.00 indicate the risk adjuster model under-predicted the expenditures of a given group -- i.e., a predictive ratio of 0.90 indicates that the risk adjuster model under-predicted the expenditures of that group by 10 percent. Similarly, a predictive ratio above 1.00 for a risk adjuster model for a given group indicates that the risk adjuster model over-predicted the expenditures of that group. Thus, bars that extend below the horizontal axis of 1.00 indicate groups for whom a risk adjuster model under-predicted their expenditures, and bars extending above the horizontal axis indicate groups for whom a risk adjuster model over-predicted their expenditures.

CHART TWO
DISTRIBUTION OF PREDICTIVE RATIOS FOR RANDOM SAMPLES OF SIZE 5,000



Note: A "perfect" predictive ratio of 1.00 indicates that the risk adjuster model exactly predicted the expenditures of a given group. Predictive ratios of less than 1.00 indicate the risk adjuster model under-predicted the expenditures of a given group -- i.e., a predictive ratio of 0.90 indicates that the risk adjuster model under-predicted the expenditures of that group by 10 percent. Similarly, a predictive ratio above 1.00 for a risk adjuster model for a given group indicates that the risk adjuster model over-predicted the expenditures of that group. Thus, bars that extend below the horizontal axis of 1.00 indicate groups for whom a risk adjuster model under-predicted their expenditures, and bars extending above the horizontal axis indicate groups for whom a risk adjuster model over-predicted their expenditures.

CHART THREE
DISTRIBUTION OF PREDICTIVE RATIOS FOR RANDOM SAMPLES OF SIZE 50,000



Note: A "perfect" predictive ratio of 1.00 indicates that the risk adjuster model exactly predicted the expenditures of a given group. Predictive ratios of less than 1.00 indicate the risk adjuster model under-predicted the expenditures of a given group -- i.e., a predictive ratio of 0.90 indicates that the risk adjuster model under-predicted the expenditures of that group by 10 percent. Similarly, a predictive ratio above 1.00 for a risk adjuster model for a given group indicates that the risk adjuster model over-predicted the expenditures of that group. Thus, bars that extend below the horizontal axis of 1.00 indicate groups for whom a risk adjuster model under-predicted their expenditures, and bars extending above the horizontal axis indicate groups for whom a risk adjuster model over-predicted their expenditures.

models, particularly for the ADG-Hosdom model, than for the "AAPCC" model. In addition, the range of predictive ratios narrows as the size of the random samples increases -- both the standard deviation of the distribution of predictive ratios and the ranges (i.e., minimum to maximum and inter-quartile ranges) become smaller as the sample size increases. Both JHU risk adjuster models do a relatively good job at forecasting the expenditures of larger groups -- that is, groups of 5,000 individuals or more.

The performance of the ADG-MDC and ADG-Hosdom risk adjuster models is quite similar across the repeated random subsamples. Median and mean predictive ratios for both models are almost identical for all random sample sizes, as are the standard deviations of these predictive ratio distributions. In addition, the minimum and maximum and 5th, 10th, 25th, 75th, 90th, and 95th percentiles of the predictive ratio distributions for the ADG-MDC and ADG-Hosdom models are also virtually identical. In addition, the range of predictive ratios across the various group sizes is smaller for the ADG-MDC and ADG-Hosdom models than for the "AAPPC."

On the other hand, the range of predictive ratios is fairly broad, even for quite large random groups. For example, approximately five percent (i.e., the fifth and 95th percentiles) of the random groups of size 50,000 have predictive ratios typically less than 0.90 or greater than 1.10 for both the ADG-MDC and ADG-Hosdom models. This suggests that plans with 50,000 enrollees enrolled at random have greater than a five percent chance of incurring losses or gains of 10 percent or more if payments were based on either the ADG-MDC or the ADG-Hosdom model, solely due to random chance.

b. Risk Adjusters with Reinsurance

The previous findings indicate that even for large random groups (i.e., 50,000 individuals), risk adjuster models alone often can over-predict or under-predict group expenditures for a small percentage of these random groups by over 10 percent. It is possible that the tendency of risk adjuster models to over-predict or under-predict the expenditures for large groups could be improved by incorporating reinsurance.

We used the truncated estimates described in ***Chapter III*** to simulate the effects of stop-loss reinsurance. In these estimates, JHU reestimated the parameters of each risk adjuster model after truncating expenditures at either $50,000 or $100,000. Using these results, we have simulated the effects of stop-loss reinsurance with stop-loss thresholds of $50,000 and $100,000 for each risk adjuster model (i.e., the ADG-MDC, ADG-Hosdom, and "AAPCC" models). Above these thresholds, we have assumed that the reinsurance system would pay 80 percent of all costs, leaving the plans responsible for a 20 percent coinsurance above the reinsurance thresholds. The effects of reinsurance on the distribution of predictive ratios for very large (50,000 individuals) random groups for each model were then compared (***Table IV-2***).

As expected, incorporating retrospective reinsurance reduced the range of predictive ratios for all three models for random groups of 50,000 individuals. The total range, inter-quartile ranges, and standard deviation of the distributions declined for all three risk adjuster models after reinsurance is included. The reduction in these ranges, however, was not large. For example, the total range for the ADG-MDC model declined from 0.90 -- 1.13 (0.25) without

TABLE IV-2: DISTRIBUTION OF PREDICTIVE RATIOS FOR REPEATED RANDOM SAMPLES OF 50,000 WITH REINSURANCE											
Model	Minimum	5th Percentile	10th Percentile	25th Percentile	Median	75th Percentile	90th Percentile	95th Percentile	Maximum	Mean	Std. Dev.
"AAPCC"											
No Reinsurance	0.8866	0.8901	0.8923	0.9040	1.0040	1.0461	1.1086	1.1325	1.1607	0.9994	0.0777
$100,000	0.8914	0.8962	0.8986	0.9106	0.9962	1.0394	1.0999	1.1239	1.1494	0.9970	0.0727
$50,000	0.8935	0.8989	0.9017	0.9124	0.9867	1.0273	1.0836	1.1061	1.1294	0.9899	0.0660
ADG-MDC											
No Reinsurance	0.8999	0.9063	0.9091	0.9344	0.9972	1.0545	1.0877	1.1127	1.1294	0.9990	0.0665
$100,000	0.9075	0.9125	0.9153	0.9377	0.9916	1.0486	1.0800	1.1047	1.1193	0.9966	0.0618
$50,000	0.9123	0.9157	0.9177	0.9357	0.9838	1.0350	1.0669	1.0881	1.1019	0.9897	0.0555
ADG-Hosdom											
No Reinsurance	0.9072	0.9129	0.9144	0.9363	1.0002	1.0410	1.0852	1.1139	1.1360	0.9998	0.0638
$100,000	0.9144	0.9185	0.9202	0.9392	0.9937	1.0343	1.0797	1.1055	1.1255	0.9974	0.0592
$50,000	0.9175	0.9201	0.9224	0.9369	0.9282	1.0235	1.0667	1.0891	1.1070	0.9903	0.0533

Note: A "perfect" predictive ratio of 1.00 indicates that the risk adjuster model exactly predicted the expenditures of a given group. Predictive ratios of less than 1.00 indicate the risk adjuster model under-predicted the expenditures of a given group -- i.e., a predictive ratio of 0.90 indicates that the risk adjuster model under-predicted the expenditures of that group by 10 percent. Similarly, a predictive ratio above 1.00 for a risk adjuster model for a given group indicates that the risk adjuster model over-predicted the expenditures of that group.

reinsurance to only 0.91 to 1.12 (0.21) with a reinsurance threshold of $100,000 and to 0.91 to 1.10 (0.19) for a reinsurance threshold of $50,000. The effects of reinsurance on the predictive ratios of the ADG-Hosdom and "AAPCC" risk adjuster models is similar.[2] These findings suggest that with random groups relatively few individuals reflect expenditures above $100,000.

2. NON-RANDOM GROUPS

Before discussing these results, we explain the content of the detailed tables presented below. Each table compares the predictive ratios and adjusted R square statistics for the three models: 1) "AAPCC"; 2) ADG-MDC; and 3) ADG-Hosdom. These results are presented in the "Adj. R Square" and "Ratio" columns, respectively. We also present the size of each group in the "N" column. Finally, we present the actual mean 1992 expenditures for each group (the "Actual" column) and the mean expenditures for that group predicted by each model (the "Predicted" columns).

a. Age, Gender, and Race

Table IV-3 includes predictive ratios and adjusted R square statistics for different age/gender and racial groups for the "AAPCC", ADG-MDC, and ADG-Hosdom models. ***Charts IV-4 through IV-6*** replicate the predictive ratio results.

All three models perform quite well across age/gender groups. The "AAPCC" predictive ratios vary from 0.93 to 1.09, while these ratios vary from 0.95 to 1.04 for the ADG-MDC model and from 0.96 to 1.02 for the ADG-Hosdom model. The range in predictive ratios across these age/gender groups for both the ADG-MDC and ADG-Hosdom models are narrower than the range for the "AAPCC," indicating that both JHU models better adjust for differences in relative risk by age and gender. For some individual age/gender groups, however, the predictive ratios for the "AAPCC" are closer to 1.00 than either the ADG-MDC and ADG-Hosdom models. For example, the "AAPCC" performs slightly better than the two JHU models for males age 85 and above.

An analysis of the adjusted R square statistics indicated that they were much higher for the age and gender groups for the two JHU risk adjuster models than for the "AAPCC" model. In addition, the adjusted R square statistics also tended to be higher across groups for the ADG-MDC than for the ADG-Hosdom model. These higher adjusted R square statistics suggest that it would be harder to "cream skim" the better risks within these age/gender and racial groups for the two JHU risk adjuster models .

At the same time, it is worth noting important divergences between the predictive ratio and adjusted R square results. For example, the adjusted R square statistics for the ADG-MDC model are greater than those for the ADG-Hosdom model for every age/gender group, but the predictive ratios for the ADG-Hosdom model are closer to 1.00 than those of the ADG-MDC model in eight out of ten cases. Better individual predictive accuracy thus does not imply better prediction at the group level.

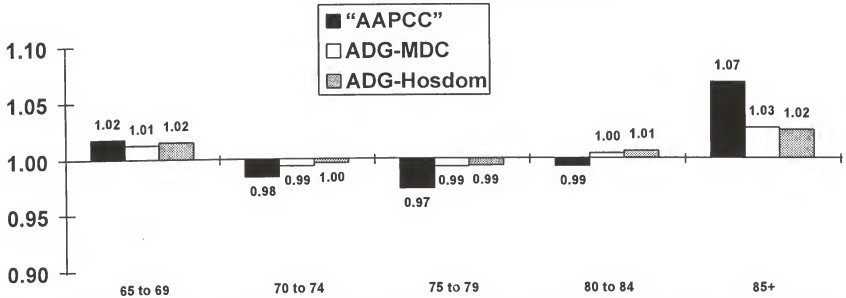
[2] The reinsurance schemes are not revenue neutral. This is because the payments above the reinsurance thresholds are only 80 percent, not 100 percent. It is unlikely, however, that a reinsurance scheme with reimbursement of costs above the threshold at a payment rate of 100 percent would be adopted, because such a scheme would provide plans with no incentives to contain the costs of extremely high cost enrollees.

TABLE IV-3: PREDICTIVE RATIOS AND ADJUSTED R SQUARE STATISTICS BY AGE/GENDER AND RACE												
			"AAPCC"			ADG-MDC			ADG-Hosdom			
	N	Actual	Predicted	Adj. R. Square a/	Ratio	Predicted	Adj. R. Square a/	Ratio	Predicted	Adj. R. Square a/	Ratio	
All												
	620,507	$ 3,498	$ 3,503	0.97%	1.0014	$ 3,499	5.96%	1.0003	$ 3,502	5.30%	1.0012	
Age/Gender												
Female 65 to 69	87,876	$ 2,357	$ 2,398	1.22%	1.0178	$ 2,384	7.89%	1.0117	$ 2,392	6.62%	1.0150	
Female 70 to 74	99,633	$ 2,924	$ 2,875	0.60%	0.9832	$ 2,906	6.12%	0.9937	$ 2,912	5.37%	0.9959	
Female 75 to 79	80,340	$ 3,524	$ 3,428	0.33%	0.9728	$ 3,499	5.94%	0.9928	$ 3,500	5.17%	0.9933	
Female 80 to 84	56,861	$ 3,965	$ 3,936	0.23%	0.9926	$ 3,982	5.39%	1.0042	$ 3,987	4.73%	1.0055	
Female 85+	52,291	$ 4,442	$ 4,741	0.18%	1.0673	$ 4,560	4.46%	1.0266	$ 4,552	3.68%	1.0247	
Male 65 to 69	69,367	$ 2,898	$ 3,169	0.55%	1.0937	$ 3,001	5.80%	1.0357	$ 2,995	5.11%	1.0338	
Male 70 to 74	73,686	$ 3,572	$ 3,615	0.35%	1.0120	$ 3,580	5.10%	1.0023	$ 3,580	4.40%	1.0023	
Male 75 to 79	51,740	$ 4,437	$ 4,130	0.22%	0.9307	$ 4,222	4.66%	0.9516	$ 4,239	4.07%	0.9554	
Male 80 to 84	29,748	$ 4,875	$ 4,609	0.12%	0.9455	$ 4,788	4.71%	0.9821	$ 4,798	4.35%	0.9842	
Male 85+	18,965	$ 5,304	$ 5,297	0.08%	0.9986	$ 5,356	3.93%	1.0099	$ 5,344	3.60%	1.0075	
Race												
White	555,200	$ 3,484	$ 3,482	0.96%	0.9993	$ 3,504	5.82%	1.0058	$ 3,427	5.18%	0.9836	
Black	45,625	$ 3,882	$ 3,826	0.87%	0.9855	$ 3,626	6.91%	0.9340	$ 3,528	6.17%	0.9088	
Other	19,682	$ 3,007	$ 3,354	1.17%	1.1156	$ 3,068	9.08%	1.0204	$ 2,946	7.26%	0.9797	

Note: A "perfect" predictive ratio of 1.00 indicates that the risk adjuster model exactly predicted the expenditures of a given group. Predictive ratios of less than 1.00 indicate the risk adjuster model under-predicted the expenditures of a given group -- i.e., a predictive ratio of 0.90 indicates that the risk adjuster model under-predicted the expenditures of that group by 10 percent. Similarly, a predictive ratio above 1.00 for a risk adjuster model for a given group indicates that the risk adjuster model over-predicted the expenditures of that group.

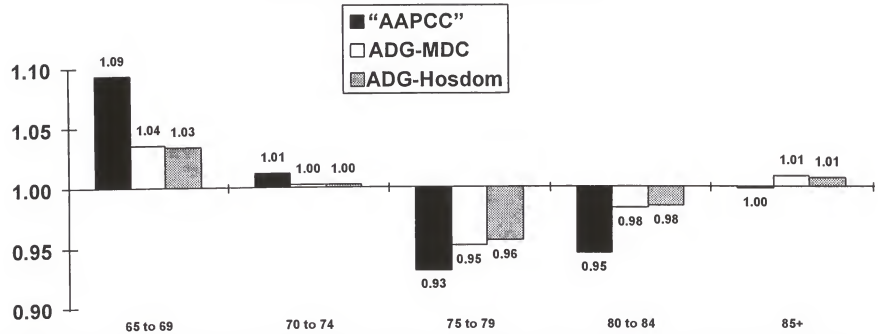
a/ The adjusted r-square statistics reported in this table were calculated by reestimating each risk adjustor model for each non-random group selected from the test split-half sample. In contrast, the predictive ratios presented in the table were calculated by assuming estimates for each model that were calculated using the entire development split-half sample to each of the non-random groups selected from the test half-sample.

CHART FOUR
PREDICTIVE RATIOS FOR FEMALE AGE GROUPS



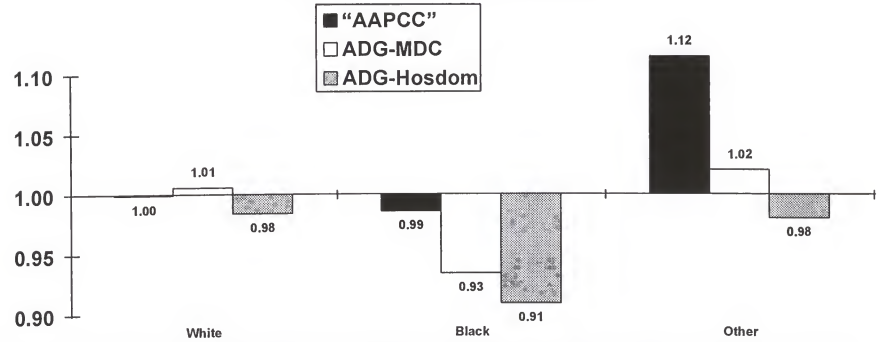
Note: A "perfect" predictive ratio of 1.00 indicates that the risk adjuster model exactly predicted the expenditures of a given group. Predictive ratios of less than 1.00 indicate the risk adjuster model under-predicted the expenditures of a given group -- i.e., a predictive ratio of 0.90 indicates that the risk adjuster model under-predicted the expenditures of that group by 10 percent. Similarly, a predictive ratio above 1.00 for a risk adjuster model for a given group indicates that the risk adjuster model over-predicted the expenditures of that group. Thus, bars that extend below the horizontal axis of 1.00 indicate groups for whom a risk adjuster model under-predicted their expenditures, and bars extending above the horizontal axis indicate groups for whom a risk adjuster model over-predicted their expenditures.

CHART FIVE
PREDICTIVE RATIOS FOR MALE AGE GROUPS



Note: A "perfect" predictive ratio of 1.00 indicates that the risk adjuster model exactly predicted the expenditures of a given group. Predictive ratios of less than 1.00 indicate the risk adjuster model under-predicted the expenditures of a given group -- i.e., a predictive ratio of 0.90 indicates that the risk adjuster model under-predicted the expenditures of that group by 10 percent. Similarly, a predictive ratio above 1.00 for a risk adjuster model for a given group indicates that the risk adjuster model over-predicted the expenditures of that group. Thus, bars that extend below the horizontal axis of 1.00 indicate groups for whom a risk adjuster model under-predicted their expenditures, and bars extending above the horizontal axis indicate groups for whom a risk adjuster model over-predicted their expenditures.

CHART SIX
PREDICTIVE RATIOS FOR RACIAL GROUPS



Note: A "perfect" predictive ratio of 1.00 indicates that the risk adjuster model exactly predicted the expenditures of a given group. Predictive ratios of less than 1.00 indicate the risk adjuster model under-predicted the expenditures of a given group -- i.e., a predictive ratio of 0.90 indicates that the risk adjuster model under-predicted the expenditures of that group by 10 percent. Similarly, a predictive ratio above 1.00 for a risk adjuster model for a given group indicates that the risk adjuster model over-predicted the expenditures of that group. Thus, bars that extend below the horizontal axis of 1.00 indicate groups for whom a risk adjuster model under-predicted their expenditures, and bars extending above the horizontal axis indicate groups for whom a risk adjuster model over-predicted their expenditures.

b. The Use of Medical Services

Using data on medical service utilization in 1991, the following groups were created: (1) individuals with no inpatient admissions; (2) individuals with one inpatient admission; (3) individuals with two inpatient admissions; (4) individuals with three or more inpatient admissions; and (5) individuals with no inpatient admissions but who were heavy users of physician services (as measured by physician RVUs). These results are presented in ***Table IV-4***.

The "AAPCC" model performed by far the worst for all four groups defined by the number of inpatient admissions in 1991. For each of these four groups, the adjusted R square statistics for the "AAPCC" model were much lower than those for the ADG-MDC and ADG-Hosdom models, indicating that the "AAPCC" accounted for less of the individual variation in medical expenditures. In addition, the "AAPCC" also performed worse on a group level than did either JHU risk adjuster model, as measured by the predictive ratios for these four groups.

The ADG-MDC model, however, performed noticeably better than the ADG-Hosdom model for groups with multiple hospital admissions. The predictive ratios for the ADG-MDC model for the group with two inpatient admissions was 1.01 and 0.97 for the group with three or more admissions, while the corresponding predictive ratios for the ADG-Hosdom model were 0.91 and 0.66, respectively.

This result is not unexpected. Individuals with hospital admissions in one year typically have higher medical expenditures in the following year. The ADG-MDC model explicitly uses information on the number and type of hospital admissions in year one (the MDC variables) to help predict medical expenditures in year two.

All three risk adjuster models had difficulty predicting the 1992 expenditures of individuals who were high users of physician services in 1991 but who were not admitted to the hospital. The predictive ratios of the ADG-MDC (0.81) and ADG-Hosdom (0.80) models for this group diverged considerably from 1.00, but were both much closer than the predictive ratio for the "AAPCC" model (0.53). Similar results for these medical service groups are presented in ***Appendix IV-1***.

c. Medical Conditions

Both JHU risk adjuster models and the "AAPCC" were also evaluated for groups of 17 different medical conditions in 1991 that were of particular interest to clinicians at HCFA-ORD. These conditions are:

- depression;
- alcohol and drug abuse
- hypertensive heart/renal disease;
- benign/unspecified hypertension;
- diabetes with complications;
- diabetes without complications
- heart failure/cardiomyopathy;
- acute myocardial infarction;
- other heart disease;
- chronic obstructive pulmonary disease;

TABLE IV-4: PREDICTIVE RATIOS AND ADJUSTED R SQUARE STATISTICS BY USE OF MEDICAL SERVICES												
			"AAPCC"			ADG-MDC			ADG-Hosdom			
	N	Actual	Predicted	Adj. R. Square a/	Ratio	Predicted	Adj. R. Square a/	Ratio	Predicted	Adj. R. Square a/	Ratio	
Number of Hospital Discharges in 1991												
None	507,919	$ 2,743	$ 3,442	0.78%	1.2548	$ 2,720	2.49%	0.9918	$ 2,834	2.74%	1.0332	
One	76,856	$ 5,710	$ 3,750	0.39%	0.6568	$ 5,913	1.78%	1.0356	$ 6,042	1.84%	1.0582	
Two	22,894	$ 8,419	$ 3,858	0.29%	0.4583	$ 8,527	1.62%	1.0128	$ 7,638	1.56%	0.9072	
Three or More	12,838	$ 13,427	$ 3,938	0.59%	0.2933	$ 12,962	4.72%	0.9654	$ 8,897	2.52%	0.6626	
High Users of Physician Services (1,527 RVUs or More in 1991) with No Hospital Admissions in 1991												
High Users	54,162	$ 6,816	$ 3,621	0.61%	0.5313	$ 5,511	5.73%	0.8086	$ 5,548	4.71%	0.8008	

Note: A "perfect" predictive ratio of 1.00 indicates that the risk adjuster model exactly predicted the expenditures of a given group. Predictive ratios of less than 1.00 indicate the risk adjuster model under-predicted the expenditures of a given group -- i.e., a predictive ratio of 0.90 indicates that the risk adjuster model under-predicted the expenditures of that group by 10 percent. Similarly, a predictive ratio above 1.00 for a risk adjuster model for a given group indicates that the risk adjuster model over-predicted the expenditures of that group.

a/ The adjusted r-square statistics reported in this table were calculated by reestimating each risk adjustor model for each non-random group selected from the test split-half sample. In contrast, the predictive ratios presented in the table were calculated by assuming estimates for each model that were calculated using the entire development split-half sample to each of the non-random groups selected from the test half-sample.

- colorectal cancer;
- breast cancer;
- lung/pancreas cancer;
- other stroke;
- intracerebral hemorrhage;
- hip fracture; and
- arthritis.

These groups were selected by HCFA clinicians for this project. Chronic conditions are of particular interest, because individuals with these conditions may currently either be selected out of HMOs or be underserved if enrolled. The ICD-9-CM codes used to specify these groups are presented in ***Appendix IV-2.***

If a given risk adjuster model is able to predict the expenditures of individuals with these conditions accurately, payment systems using the risk adjuster system should improve access to HMOs for beneficiaries with chronic diseases and medical conditions and should promote the provision of appropriate care for those that are enrolled. These results are presented in ***Table IV-5***.

For individuals with one of these conditions in 1991, both the ADG-MDC and ADG-Hosdom models did a far better job of predicting medical expenditures in 1992. The "AAPCC" did the best job at predicting the 1992 medical expenditures for only two condition groups, diabetes without complications and breast cancer. In all other cases, either the ADG-MDC or ADG-Hosdom model was the best performer. The ADG-MDC model was the best at predicting the expenditures for nine medical condition groups (again, defined with the condition in 1991), while the ADG-Hosdom performed the best for the remaining five groups.

d. Expenditure Groups

One of the most important features of a well-functioning risk adjuster system is its ability to allow plans and providers to enroll individuals who have been heavy users of services without financial penalty. Risk adjuster systems that do not adjust for the higher risks of enrollees who have been heavy users of services and that tend to overcompensate plans for enrollees who have used fewer medical services in the past will encourage favorable selection into capitated payment plans. For example, under the AAPCC, plans have strong incentives to target their enrollment towards healthier individuals.

Table IV-6 includes the predictive ratios and adjusted R square statistics for expenditure quintile groups in 1991. The results for 1991 expenditures groups are also reproduced in ***Chart Seven***. Similar results for expenditure quintile groups in 1992 are presented in ***Appendix IV-1***.

Chart Seven indicates that the predictive ratios are quite high for the first three 1991 expenditure quintiles and quite low for the remaining top two expenditure quintiles for the "AAPCC" model. For example, the ratio of 2.34 for the first 1991 expenditure quintile

TABLE IV-5: PREDICTIVE RATIOS AND ADJUSTED R SQUARE STATISTICS BY CLINICAL CATEGORIES IN 1991

			"AAPCC"			ADG-MDC			ADG-Hosdom		
1991 Conditions	N	Actual	Predicted	Adj. R. Square a/	Ratio	Predicted	Adj. R. Square a/	Ratio	Predicted	Adj. R. Square a/	Ratio
None b/	233,088	$ 2,091	$ 3,339	1.08%	1.5971	$ 2,267	3.51%	1.0844	$ 2,211	3.20%	1.0575
Depression	2,329	$ 3,575	$ 3,373	0.76%	0.9437	$ 3,546	0.45%	0.9921	$ 3,652	4.69%	1.0215
Alcohol /Drug Abuse	279	$ 4,685	$ 3,710	-0.13%	0.7918	$ 5,214	-6.12%	1.1128	$ 5,667	-3.20%	1.2096
Hypertensive Heart/Renal Disease	6,244	$ 2,891	$ 3,385	0.77%	1.1712	$ 3,372	3.54%	1.1664	$ 3,495	2.57%	1.2091
Benign/Unspecified Hypertension	74,359	$ 2,450	$ 3,320	0.62%	1.3546	$ 2,589	2.51%	1.0564	$ 2,608	2.23%	1.0643
Diabetes with Complications	1,879	$ 3,958	$ 3,505	0.96%	0.8854	$ 4,078	3.83%	1.0301	$ 4,192	2.65%	1.0591
Diabetes without Complications	28,789	$ 3,723	$ 3,448	0.63%	0.9260	$ 3,176	3.44%	0.8528	$ 3,210	3.03%	0.8621
Heart Failure/Cardiomyopathy	9,162	$ 5,743	$ 4,097	0.17%	0.7133	$ 5,149	3.80%	0.8965	$ 5,060	3.16%	0.8810
Acute Myocardial Infarction	1,034	$ 5,836	$ 3,697	0.09%	0.6335	$ 5,151	1.92%	0.8827	$ 5,877	1.98%	1.0071
Other Heart Disease	86,004	$ 4,679	$ 3,684	0.53%	0.7873	$ 4,844	3.37%	1.0353	$ 4,844	2.78%	1.0354
Chronic Obstructive Pulmonary Disease	55,425	$ 5,247	$ 3,586	0.86%	0.6834	$ 4,940	5.83%	0.9415	$ 4,847	4.59%	0.9238
Colorectal Cancer	4,898	$ 6,692	$ 3,602	0.30%	0.5383	$ 5,845	5.42%	0.8734	$ 6,010	3.95%	0.8981
Breast Cancer	7,867	$ 3,339	$ 3,095	0.54%	0.9270	$ 4,738	6.22%	1.4189	$ 4,749	5.16%	1.4223
Lung/Pancreas Cancer	3,055	$ 10,508	$ 3,531	3.10%	0.3360	$ 7,513	4.93%	0.7150	$ 6,924	3.97%	0.6589
Other Stroke	17,982	$ 6,839	$ 3,856	0.49%	0.5638	$ 6,398	4.91%	0.9355	$ 6,777	4.19%	0.9911
Intracerebral Hemorrhage	570	$ 8,512	$ 3,757	-0.53%	0.4415	$ 6,904	-1.54%	0.8111	$ 7,833	-0.11%	0.9203
Hip Fracture	4,981	$ 6,455	$ 4,212	0.16%	0.6525	$ 6,265	3.63%	0.9704	$ 6,798	2.68%	1.0531
Arthritis	82,562	$ 4,420	$ 3,603	0.80%	0.8151	$ 4,231	5.15%	0.9572	$ 4,320	4.57%	0.9773

Note: A "perfect" predictive ratio of 1.00 indicates that the risk adjuster model exactly predicted the expenditures of a given group. Predictive ratios of less than 1.00 indicate the risk adjuster model under-predicted the expenditures of a given group -- i.e., a predictive ratio of 0.90 indicates that the risk adjuster model under-predicted the expenditures of that group by 10 percent. Similarly, a predictive ratio above 1.00 for a risk adjuster model for a given group indicates that the risk adjuster model over-predicted the expenditures of that group. Thus, bars that extend below the horizontal axis of 1.00 indicate groups for whom a risk adjuster model under-predicted their expenditures, and bars extending above the horizontal axis indicate groups for whom a risk adjuster model over-predicted their expenditures. The adjusted R square statistics associated with some chronic conditions for some of the risk adjuster models are negative, a rare, though possible occurrence.

a/ The adjusted r-square statistics reported in this table were calculated by reestimating each risk adjustor model for each non-random group selected from the test split-half sample. In contrast, the predictive ratios presented in the table were calculated by assuming estimates for each model that were calculated using the entire development split-half sample to each of the non-random groups selected from the test half-sample.

b/ "None" refers to all patients not grouped into any of the 17 disease categories.

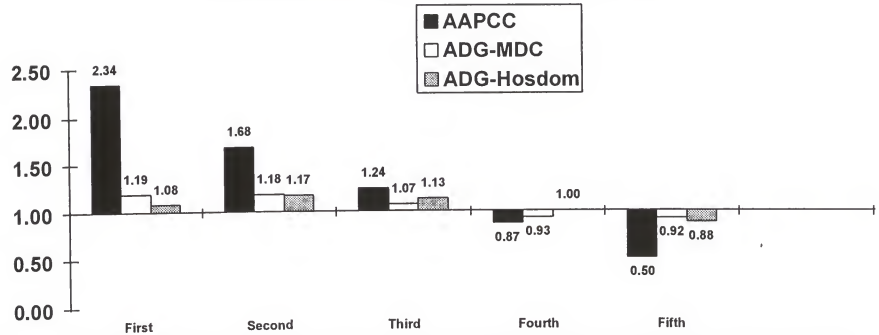
TABLE IV-6: PREDICTIVE RATIOS AND ADJUSTED R SQUARE STATISTICS BY EXPENDITURE GROUPS IN 1991												
			"AAPCC"			ADG-MDC			ADG-Hosdom			
	N	Actual	Predicted	Adj. R. Square a/	Ratio	Predicted	Adj. R. Square a/	Ratio	Predicted	Adj. R. Square a/	Ratio	
Expenditure Quintiles in 1991												
First	116,325	$ 1,415	$ 3,315	0.53%	2.3417	$ 1,686	0.65%	1.1913	$ 1,525	0.66%	1.0777	
Second	128,165	$ 2,007	$ 3,375	0.73%	1.6819	$ 2,365	1.00%	1.1788	$ 2,355	1.00%	1.1736	
Third	130,926	$ 2,807	$ 3,475	0.75%	1.2379	$ 3,002	1.05%	1.0693	$ 3,160	1.07%	1.1258	
Fourth	130,449	$ 4,132	$ 3,581	0.60%	0.8667	$ 3,842	1.30%	0.9297	$ 4,140	1.35%	1.0019	
Fifth	114,642	$ 7,569	$ 3,795	0.45%	0.5014	$ 6,972	3.45%	0.9212	$ 6,630	2.60%	0.8759	

Note: A "perfect" predictive ratio of 1.00 indicates that the risk adjuster model exactly predicted the expenditures of a given group. Predictive ratios of less than 1.00 indicate the risk adjuster model under-predicted the expenditures of a given group -- i.e., a predictive ratio of 0.90 indicates that the risk adjuster model under-predicted the expenditures of that group by 10 percent. Similarly, a predictive ratio above 1.00 for a risk adjuster model for a given group indicates that the risk adjuster model over-predicted the expenditures of that group.

The size of the expenditure groups was not the same, as one might expect. Some of the groups were larger, because an unusual number of individuals had medical expenditures equal to the "break points" between the quintiles. We assigned individuals with medical expenditures equal to these "break point" levels to the lower of the two quintile groups.

a/ The adjusted r-square statistics reported in this table were calculated by reestimating each risk adjustor model for each non-random group selected from the test split-half sample. In contrast, the predictive ratios presented in the table were calculated by assuming estimates for each model that were calculated using the entire development split-half sample to each of the non-random groups selected from the test half-sample.

CHART SEVEN
PREDICTIVE RATIOS FOR 1991 EXPENDITURE QUINTILE GROUPS



Note: A "perfect" predictive ratio of 1.00 indicates that the risk adjuster model exactly predicted the expenditures of a given group. Predictive ratios of less than 1.00 indicate the risk adjuster model under-predicted the expenditures of a given group -- i.e., a predictive ratio of 0.90 indicates that the risk adjuster model under-predicted the expenditures of that group by 10 percent. Similarly, a predictive ratio above 1.00 for a risk adjuster model for a given group indicates that the risk adjuster model over-predicted the expenditures of that group. Thus, bars that extend below the horizontal axis of 1.00 indicate groups for whom a risk adjuster model under-predicted their expenditures, and bars extending above the horizontal axis indicate groups for whom a risk adjuster model over-predicted their expenditures.

indicates that the "AAPCC" model over-predicts the 1992 expenditures for this group by 134 percent. Plans receiving payments that enroll individuals with low medical expenditures in 1991 stand to gain handsomely in a payment system based on the "AAPCC." Conversely, the predictive ratio of 0.50 for the fifth 1991 expenditure quintile for the "AAPPC" indicates that plans on average would receive only half the 1992 costs of caring for individuals with high costs in 1991. Thus, the "AAPPC" provides strong incentives for plans to identify and avoid enrolling individuals with high medical costs in the preceding year.

The large differences in predictive ratios across 1991 expenditure quintiles for the "AAPCC" did not occur for either the ADG-MDC or ADG-Hosdom models. Unlike the "AAPCC" range in predictive ratios that extended from 0.50 for the fifth expenditure quintile in 1991 to 2.34 for the first, these ranges narrowed to 0.92 to 1.19 for the ADG-MDC model and from 0.88 to 1.17 for the ADG-Hosdom model.[3] While plans would still benefit from enrolling individuals with low medical expenditures in the past and would suffer if they enroll individuals with high previous medical expenditures if payments were based on either JHU Medicare risk adjustment model, the incentives for favorable selection and the avoidance of high cost individuals are much weaker than those under the "AAPCC".

e. Geographic Region

Finally, ***Table IV-7*** includes predictive ratios and adjusted R square statistics for the nine census division regions for each of the three risk adjuster models. The dependent variable used in these models, actual Medicare expenditures for services in 1992, included local price variation. The resulting model parameters (the estimated coefficients) thus reflected a national average that did not reflect a specific regions' conditions. The predictive ratios for census division regions for a given risk adjuster model thus differed from 1.00 for several reasons. First, as with other non-random groups, a given risk adjuster model may not have estimated the costs for a given region that well. Second, the predictive ratios for census division regions may have also reflected differing practice patterns across regions. Third, it was also possible that regional variations in coding practices could have affected our results. Finally, the predictive ratios for census division regions could also incorporate differences in medical prices across the regions.

For each census division region, all three risk adjuster models either over-predicted (predictive ratio > 1.00) or under-predicted (predictive ratio < 1.00) that region's actual mean expenditures. For example, the "AAPCC," ADG-MDC, and ADG-Hosdom models over-predicted expenditures in the East North Central, West North Central, West South Central, and Mountain regions, and under-predicted expenditures in the New England, Middle Atlantic, and Pacific regions.

For the most part, both the ADG-MDC and ADG-Hosdom models better predicted medical expenditures by region than the "AAPCC." The predictive ratios for the ADG-MDC model by region ranged from 0.90 to 1.15, and from 0.90 to 1.15 for the ADG-Hosdom model, but ranges from 0.88 to 1.18 for the "AAPCC" model. For three regions (the South Atlantic, East South Central, and the Pacific regions), however, the "AAPCC" outperformed both JHU risk adjuster models. In addition, the adjusted R square statistics for the ADG-MDC and ADG-Hosdom models were consistently much larger than those of the "AAPCC" model in all nine census division regions.

[3] The highest predictive ratio for the ADG-Hosdom model is for the second expenditure quintile group in 1991.

TABLE IV-7: PREDICTIVE RATIOS AND ADJUSTED R SQUARE STATISTICS BY REGION												
			"AAPCC"			ADG-MDC			ADG-Hosdom			
	N	Actual	Predicted	Adj. R. Square a/	Ratio	Predicted	Adj. R. Square a/	Ratio	Predicted	Adj. R. Square a/	Ratio	
Census Division Region: 1991												
New England	36,976	$ 3,647	$ 3,485	1.25%	0.9557	$ 3,499	7.02%	0.9593	$ 3,486	6.30%	0.9557	
Middle Atlantic	105,428	$ 3,942	$ 3,460	0.91%	0.8777	$ 3,532	5.18%	0.8961	$ 3,536	4.83%	0.8970	
South Atlantic	112,087	$ 3,437	$ 3,441	0.98%	1.0010	$ 3,456	6.41%	1.0053	$ 3,455	5.57%	1.0050	
East North Central	49,169	$ 2,975	$ 3,510	0.72%	1.1797	$ 3,426	5.48%	1.1515	$ 3,426	5.18%	1.1516	
East South Central	116,328	$ 3,462	$ 3,505	1.17%	1.0123	$ 3,547	6.29%	1.0247	$ 3,578	5.52%	1.0335	
West North Central	41,118	$ 3,241	$ 3,620	0.98%	1.1171	$ 3,670	7.05%	1.1323	$ 3,592	6.01%	1.1084	
West South Central	62,570	$ 3,362	$ 3,543	1.00%	1.0539	$ 3,493	7.13%	1.0389	$ 3,457	6.20%	1.0284	
Mountain	30,773	$ 2,949	$ 3,450	0.76%	1.1700	$ 3,378	6.02%	1.1454	$ 3,401	5.15%	1.1535	
Pacific	70,272	$ 3,727	$ 3,600	1.10%	0.9661	$ 3,485	5.81%	0.9351	$ 3,537	4.90%	0.9491	
Unknown	1,236	$ 983	$ 2,874	1.52%	2.9246	$ 1,302	1.61%	1.3248	$ 1,112	2.11%	1.1318	

Note: A "perfect" predictive ratio of 1.00 indicates that the risk adjuster model exactly predicted the expenditures of a given group. Predictive ratios of less than 1.00 indicate the risk adjuster model under-predicted the expenditures of a given group -- i.e., a predictive ratio of 0.90 indicates that the risk adjuster model under-predicted the expenditures of that group by 10 percent. Similarly, a predictive ratio above 1.00 for a risk adjuster model for a given group indicates that the risk adjuster model over-predicted the expenditures of that group.

a/ The adjusted r-square statistics reported in this table were calculated by reestimating each risk adjustor model for each non-random group selected from the test split-half sample. In contrast, the predictive ratios presented in the table were calculated by assuming estimates for each model that were calculated using the entire development split-half sample to each of the non-random groups selected from the test half-sample.

Given the diversity of practice styles across the nation, it is not surprising that a national model did not work well at the regional level. In practice, the JHU risk adjustment models could be geographically adjusted. In fact, as HCFA tests these models through demonstrations, regional calibrations of the models likely will be developed.

C. CONCLUSIONS

This statistical evaluation reaches four broad conclusions. First, both JHU risk adjuster models generally outperformed the "AAPCC," our model constructed to approximate the method HCFA uses to set capitation rates for TEFRA HMO and CMP risk contractors. The ADG-MDC and ADG-Hosdom models had adjusted R square statistics that tended to be five to six times as large as those of the "AAPCC." For most non-random groups, both JHU risk adjuster models had predictive ratios that were nearer to 1.00 than did the "AAPCC" model. Of particular interest for policy, the JHU models did a much better job at predicting the expenditures in the payment year of groups with different expenditure levels in the year preceding payment than did the "AAPCC" model. Thus, both the ADG-MDC and ADG-Hosdom model considerably reduce the incentives plans have to "cream-skim" the Medicare population and to avoid enrolling individuals with high prior medical use, if capitated payments were based on these risk assessment/adjustment models.

Second, neither JHU model was a clear winner. The performance of both models across the random groups and most non-random groups was nearly identical in the predictive ratio tests. The one noticeable difference was that the ADG-MDC model did a better job at controlling for differences in medical expenditures in 1992 for groups with multiple (two or three or more) inpatient hospital admissions in 1991. As mentioned above, this result is expected, because the ADG-MDC model explicitly uses inpatient admissions variables to predict 1992 medical expenditures.

Finally, these results indicate some potential problems with both JHU risk adjuster models. The predictive ratios for repeated random samples, even samples of 50,000, demonstrate that a small percentage of plans could face favorable or negative risk selection relative to either risk adjuster model, at least in some years. In addition, both the ADG-MDC and ADG-Hosdom risk adjuster models under-predict the expenditures of individuals with very high expenditures in year two (1992). There may be a need to supplement the prospective JHU risk adjuster models with some form of reinsurance or a list of diagnostic carve-out conditions.

CHAPTER V
GAMING AND ADMINISTRATIVE FEASIBILITY

In many respects, the best test of a risk adjuster model is its ability to serve as the basis of a well-designed and functional capitated payment system where plans are fairly paid for varying degrees of enrollee severity. The diagnostic-based ADG-MDC and ADG-Hosdom risk adjuster models developed by this study effort are clinically cogent and predict the medical expenditures of Medicare enrollees more accurately than an "AAPCC" risk adjuster model. A well-designed and functional capitated payment system, however, needs to embody features in addition to clinical cogency and predictive accuracy.

In particular, a successful risk adjustment system also should not be easy to "game" and must not place large administrative burdens on plans, providers, or payers. "Gaming" occurs when providers engage in strategic activities whose added costs to providers are exceeded by their added benefits (additional reimbursement). The first section of this chapter addresses the issue of gaming. This section compares the relative resistance to gaming by the ADG-MDC and ADG-Hosdom models with each other and with the AAPCC.

The second then discusses three aspects of administrative feasibility. First, will either the ADG-MDC or ADG-Hosdom models place substantial new burdens on plans and providers for collecting new information? Second, how might the results of these models be used to establish a risk adjusted capitated payment system? This section concludes with some speculations on possible provider attitudes towards risk adjustment.

A. GAMING

There are several ways providers can game a risk adjusted capitated payment system. We first discuss "upcoding," where providers manipulate diagnostic information to increase payments. This is followed by a parallel discussion of offsets to upcoding. Next, the trade-off between predictive accuracy and the susceptibility of the model to "gaming" by providers is explained. This section ends by noting the possible incentives within the ADG-MDC model that may favor increases in inpatient care.

1. Opportunity for Upcoding

Each JHU model uses a regression equation to assign each individual a risk score based on the following data (1) demographic data -- the age and gender of the enrollee; (2) insurance status -- has the enrollee ever been covered by Medicare as a result of receiving Social Security Disability Insurance (prior disability), and/or is the enrollee eligible for Medicaid; and (3) individual diagnostic data (ICD-9-CM codes) and the site of service used to specify the ADG, MDC, and hosdom variables. Higher individual risk scores translate into higher payment rates for plans. Plans and providers that want to increase enrollee risk scores for payment purposes -- as opposed to appropriate clinical reporting -- must "upcode" these data elements. This, of course, is not unique to risk adjustment as similar incentives exist in fee-for-service Medicare and HCFA's Medicare Prospective Payment System.

Of these data elements, only the individual diagnostic data and perhaps to a very limited degree the insurance status variables could be affected by plans or providers. The government collects the demographic data used by the JHU risk adjuster models, as it

currently does for its AAPCC rate setting process, and it is not clear how plans could "upcode" either age or gender.

Likewise the "insurance status" variables are not particularly amenable to gaming. The first of these, prior disability status, is clearly defined by statute. Medicare DI beneficiaries must have received DI benefits for two years before becoming Medicare eligible, and must meet a strict disability standard to qualify for DI. Even if plans were successful at identifying potential DI beneficiaries and assisted them with applying for DI benefits, those plans would need to wait for two years before these beneficiaries qualified for Medicare.

It is possible that plans could affect the second insurance status variable, eligibility for Medicaid. For example, plans might identify enrollees who may be eligible for Medicaid and aid these individuals in applying for benefits. The costs of identifying these "marginal" prospective Medicaid eligibles and assisting with their application for benefits, however, may exceed any future increases in payments. Thus, risk adjuster models such as the AAPCC that use primarily demographic and insurance status as risk adjusters are largely immune from provider upcoding.

The more likely prospects for upcoding are for plans to affect the diagnostic codes of their enrollees. ***Table V-1*** replicates the coefficient point estimates from each model (the "1992" columns). The 1992 coefficients indicate the additional estimated increase in annual expected costs per member for each risk factor. For example, expected costs in 1992 under the ADG-MDC model are $604 higher for men, $1,533 for each hospital inpatient admission in MDC category one ("nervous system"), and $542 higher for patients with a visit diagnosis in ADG category three ("time limited, minor").

In ***Table V-1,*** the 1992 estimates are also converted to hypothetical 1996 payment levels. The 1996 payment levels indicate the returns to gaming base year input statistics in a future payment year. The 1996 payment levels are calculated in two steps. First, the 1992 estimated costs are inflated to 1996. One way of inflating these costs is to increase each 1992 cost estimate using the ratio of the current U.S. monthly per Capita Cost (USPCC) for aged Medicare enrollees in 1996 to that in 1992 (i.e., $440.97/$314.99 = 1.4). These inflated estimates are presented in "1996" columns.

Second, the payer (Medicare) may not decide to reimburse plans for their full expected costs. For example, HCFA currently reimburses Medicare risk contractors at 95 percent of their expected costs as predicted by the AAPPC. Entries in the "95%" columns are obtained by thus multiplying the 1996 cost estimates by 95 percent. This is a methodology HCFA could use to set national payment rates using one of the two JHU risk adjuster models.

Plans and providers that are successful in "upcoding" their enrollees in 1995 for payment in 1996 could receive the payments indicated in the 95% columns of ***Table V-1***. The potential gross returns of upcoding could be quite substantial. For example, under the ADG-MDC risk adjuster system, plans or providers would receive more than $5,000 on an average annual, national basis for each inpatient admission in the MDC 16, 17, or 25 (Blood, Immunological, Myeloproliferative Diseases, and AIDS/HIV) and MDC 26 (transplants) categories in 1996. Similarly, the extra payment for patients successfully coded into different ADGs might exceed $1,000 per enrollee per year in both models, while a single Hosdom diagnosis in the ADG-Hosdom model could increase payments to plans by more than $2,000.

Table V-1
Possible National Annual Payment Amounts in 1992 and 1996 for the Lewin-VHI/JHU Medicare Risk Adjuster Models

	ADG-MDC			ADG-Hosdom		
	1992	1996	95%	1992	1996	95%
Demographic Variables						
Intercept	$ 608	$ 851	$ 808	$ 434	$ 608	$ 577
Male	$ 604	$ 846	$ 8041	$ 613	$ 859	$ 816
Years over Age 65	$ 67	$ 94	$ 89	$ 64	$ 89	$ 85
Ever Disabled (DI)	$1,119	$1,566	$1,488	$1,176	$1,647	$1,564
Medicaid Buy-in (QMB)	$ 761	$1,066	$1,012	$ 802	$1,123	$1,067
Hospital Dominant Marker and MDCs						
Hosdom				$1,749	$2,449	$2,327
MDCs						
MDC 1 (Nervous System)	$1,533	$2,146	$2,039			
MDC 3 or 4 (Ear, Nose, Throat and Respiratory)	$3,237	$4,532	$4,305			
MDC 5 (Circulatory System)	$1,897	$2,656	$2,523			
MDC 6 (Digestive System)	$1,759	$2,462	$2,339			
MDC 7 (Hepatobilliary System and Pancreas)	$1,030	$1,443	$1,370			
MDC 8 (Musculoskeletal System and Connective Tissue)	$1,117	$1,563	$1,485			
MDC 9 (Skin, Subcutaneous Tissue and Breast)	$1,762	$2,467	$2,343			
MDC 10 (Endocrine, Nutritional, and Metabolic Diseases)	$2,938	$4,113	$3,908			
MDC 11 (Kidney and Urinary Tract)	$2,526	$3,536	$3,359			
MDC 16, 17, or 25 (Blood, Immunological, Myeloproliferative Diseases, and AIDS/HIV)	$3,875	$5,425	$5,154			
MDC 18 (Infectious and Parasitic Diseases)	$3,061	$4,285	$4,071			
MDC 19 or 20 (Mental Diseases and Alcohol/Drug Abuse)	$1,957	$2,740	$2,603			
MDC 21 or 22 (Injuries, Poisonings, and Burns)	$1,882	$2,635	$2,503			
MDC 23 or 24 (Factors Influencing Health Status and Trauma)	$1,481	$2,073	$1,970			
MDC 26 (Transplants)	$3,944	$5,521	$5,245			
ADGs						
ADG 3 (Time Limited, Major)	$ 542	$ 759	$ 721	$ 663	$ 928	$ 882
ADG 4 (Time Limited, Major, Primary Infections)	$ 734	$1,028	$ 976	$1,503	$2,104	$1,999
ADG 6 (Asthma)	$ 818	$1,145	$1,088	$1,216	$1,702	$1,617
ADG 7 (Likely to Recur, Discrete)	$ 225	$ 315	$ 299	$ 365	$ 511	$ 485
ADG 9 (Likely to Recur, Progressive)	$ 965	$1,351	$1,284	$1,696	$2,374	$2,256
ADG 11 (Chronic Medical, Unstable)	$1,346	$1,884	$1,790	$1,415	$1,981	$1,882
ADG 16 (Chronic Specialty, Unstable, Orthopedic)	$ 650	$ 910	$ 865	$ 593	$ 830	$ 789
ADG 22 (Injuries/Adverse Effects, Major)	$ 525	$ 735	$ 698	$ 462	$ 646	$ 614
ADG 23 (Psychiatric, Time Limited, Minor)	$ 698	$ 978	$ 929	$1,222	$1,711	$1,626
ADG 25 (Psychiatric, Persistent or Recurrent, Unstable)	$ 804	$1,126	$1,070	$1,088	$1,523	$1,447
ADG 27 (Signs/Symptoms, Uncertain)	$ 460	$ 644	$ 612	$ 568	$ 75	$ 755
ADG 28 (Signs/Symptoms, Major)	$ 551	$ 771	$ 732	$ 753	$1,054	$1,001
ADG 32 (Malignancy)	$1,347	$1,886	$1,792	$1,429	$2,000	$1,900

2. Offsets to Upcoding

Several offsets would substantially reduce the incentives for upcoding under the JHU risk adjuster systems. First, the relationship between rewarding diagnostic codes in one year and then receiving additional payments in the next year is quite indirect. Second, there may be a sharp initial "spike" in coding by HMOs that is justified. Many HMOs now record only the first or second diagnosis on a hospital bill. If capitated payments depended directly on diagnostic information, HMOs would likely begin to record diagnostic data more completely.

In addition, successful upcoding requires that plans perform the following tasks:

- **Identify the codes related to each ADG, MDC and Hosdom category** -- the ADG grouper in particular is complex, and plans would need time to determine what codes cause which ADGs to "switch on." This may delay upcoding. Over time, however, plans and providers could become quite adept at acquiring information of this sort, and it is likely that such information would be provided by expert consultants to plans.
- **Target marginal patients** -- a particular ICD-9-CM might "switch on" an ADG, MDC or Hosdom variable, but that variable may already have been "switched on" by some other diagnostic code for a large number of patients. Plans and providers would need to identify those enrollees "on the margin" -- i.e., those patients who would have another variable "switch on" by a diagnostic code or codes.
- **Incur the costs associated with additional diagnoses** -- for ADGs and the Hosdom variable, coding an additional diagnosis may require little additional cost if a provider can record the additional diagnosis during an existing visit. For MDCs, however, an inpatient admission must also occur.[1]

There are costs to plans in collecting and analyzing the data needed for upcoding. Offsetting these costs, however, are the other benefits to plans that collect these data, including using these diagnostic data to manage the care of their enrollees. In addition, HCFA will need to require sufficient information to administer diagnosis-based adjustment systems. This very information could provide a basis for the collection of data required to game the system.

Along with these costs to providers, HCFA is likely to audit plans or contract out auditing activities to fiscal intermediaries (FIs) or some other third party. If plans that upcode their enrollees are at risk of being caught and then fined or otherwise penalized by an enforcement system, plans would face risks if they choose to engage in upcoding activities.

The gains to individual plans from upcoding also might be offset by lower overall payment levels. For individual plans, recording additional diagnoses and other upcoding activities will increase that plan's revenues, and each plan must balance these added revenues with the costs associated with upcoding, including enforcement. The third party payer (HCFA), however, may respond to upcoding by plans by reducing overall payment rates to plans. Risk adjuster models provide third party payers with a set of relative payment weights (for the JHU models, these are the individual risk scores). Third party payers would need to convert the risk adjuster weights into payment amounts using a conversion factor. The example above in ***Table V-1*** assumed a conversion factor in 1996 of 95 percent of the inflated

[1] During an existing inpatient admission, a provider or plan could code additional diagnoses to change (upcode) the MDC associated with that admission.

1992 risk score, the current conversion factor used to convert AAPCC weights into payment amounts.

Over time, HCFA may discern that plans and providers have all identified the primary opportunities for upcoding within the risk adjuster being used. This will result in a level of "bracket creep" that is more less common to all plans. Over time, HCFA could reduce the conversion factor it uses to set risk adjusted capitated payments to plans, to adjust for bracket creep.[2] If properly designed, these adjustments to the conversion factor might result in a zero sum game for upcoding. In addition, the new diagnostic data can be used to recalibrate the payment system periodically to lower the risk scores of "upcoded" enrollees.

3. Predictive Accuracy versus Susceptibility to Gaming

There is at present an active debate whether the informational advantages of plans and providers would lead to cream skimming under any conceivable risk adjuster system, a view held by Newhouse et al. (1989). Put simply, risk adjuster models are unlikely to use all available information, such as individual medical expenditure and utilization data, to establish risk adjuster payments. At the same time, these data are potentially available to plans for use in selecting patients. Newhouse et al. argue that this informational advantage is a decisive one. The additional predictive accuracy of the data held by plans will allow these plans to cream skim the best risks within each category. If enrollment in risk capitated plans is voluntary as it is now for Medicare, capitated plans could seek to identify and attract the best risks, and the resulting cream skimming of the best risks will increase costs to the payer (Medicare).

Other analysts are not so sure. For example, van de Ven et al. (1994) argue that risk adjuster models can be developed that are sophisticated enough to counteract the informational advantages available to plans. As risk adjuster models become better able to explain the variance in medical expenditures, the ability of plans to cream skim will be reduced. In addition, there are costs associated with cream skimming -- identifying, recruiting, and retaining the best risks are all costly activities.

Whether risk adjusters generally can eliminate cream-skimming by plans and providers is an important theoretical issue. More important, however, is comparing the abilities different risk adjuster models have at limiting cream-skimming. As shown in ***Chapter IV***, it was clear that plans who can identify and enroll individuals with low costs in 1991 (i.e., individuals in the first quintile) on average will receive payments far in excess of the costs of care for the low cost enrollees in 1992 under the AAPCC. Both the ADG-MDC and ADG-Hosdom model would sharply reduce these incentives for cream-skimming low cost patients were either model used to set HMO payments by HCFA.

4. The Locus of Care

Successful risk adjuster models should account for the predictable expenditures of relatively high cost enrollees. A major source of the greater predicative accuracy of the ADG-MDC model versus the ADG-Hosdom model is the ability to predict the high expenditures associated with inpatient hospital admissions provided by the MDC variables.

In so doing, though, the ADG-MDC risk adjuster might encourage inappropriate inpatient utilization. As indicated above, plans might receive large additional payments for their enrollees for each inpatient admission in each MDC category. These payments might

[2] Alternatively HCFA could reduce the rate of increase in the payment amount one year to the next.

provide plans an incentive to provide care in an inpatient setting. One should note, however, that the additional MDC payments plans receive for these inpatient admissions typically would not offset the costs of those admissions. For example, it is clear that the costs of providing a transplant in 1995 to an enrollee will far exceed the $5,000 payment the plan would receive in 1996 (assuming the enrollee survives during 1996).

Plans may still have incentives to increase hospital admissions in year one under the ADG-MDC risk adjuster model even if the increase in payments in year two is less than the costs of those admissions. Individuals who are not admitted to the hospital presumably still need to be treated in some other care setting. It is possible that the MDC payments in the year following these inpatient admissions exceeds the *difference* in costs between providing these services in an inpatient setting and providing them in some other setting. If so, this could provide plans with an incentive to increase inpatient admissions under the ADG-MDC risk adjuster model.

For example, suppose plans would receive an MDC payment of $2,000 per inpatient admission in the year following those inpatient stays. Next, suppose the costs of providing that care in an inpatient setting is $8,000, but that this care could be provided in some other setting for $7,000. The later alternative is clearly efficient.

If the plan decides to provide that care in an inpatient setting, it will receive $2,000 in additional payments in the next year. This $2,000 increase in payments in the next year exceeds the additional costs ($1,000) the plan incurred providing the care in an inpatient setting. In this case, the plan has an incentive to provide care in the less efficient and more costly inpatient setting.

The strength of the incentives to increase hospital admissions and the financial rewards for doing so are not clear under the ADG-MDC risk adjuster model. The simple example above did not consider the numerous reasons why plans may decide to provide care in less expensive care settings. In addition, plans must wait until the following year before receiving higher payments from the risk adjuster system, and waiting imposes costs on the plans. Furthermore, plans that treat their enrollees in more expensive care settings may not continue to enroll these individuals and thus could lose any future financial rewards (i.e., additional MDC payments). Such enrollees could decide to switch to fee for service Medicare or another plan, and it is also possible that some of them will die.

As with upcoding and cream-skimming, vigorous auditing of plans accompanied with appropriate penalties for violators could discourage plans and providers from providing care in less efficient and more costly care settings. On the other hand, enforcement is expensive. These concerns, however, are reasons to favor risk adjuster systems such as the ADG-Hosdom model that do not provide direct incentives for favoring one locus of care (inpatient care) over all others.

B. ADMINISTRATIVE FEASIBILITY

A brief discussion of three aspects of the administrative feasibility of establishing a payment system using one of the JHU risk adjuster models is provided in this section. First, whether managed care organizations (MCOs) possess the data needed for these models, and if not, how much collecting these data might cost, is discussed. Second, there are a range of issues HCFA or some other payer must address when converting the JHU models into a capitated payment system. Finally, some casual impressions of the views of providers we have encountered regarding risk adjuster models are presented.

1. Data Collection Requirements

The following discussion of the informational requirements of the JHU risk adjusters for plans and providers relies heavily on a final grant report to the Physician Prospective Payment Review Commission (PPRC 94) by Fowles et al.

Most health plans and insurers now collect demographic information for each enrollee. In 1992, the Group Health Association of America (GHAA, now the American Association of Health Plans (AAHP)) conducted a survey that indicated only 11 percent of independent practice association (IPA) and 26 percent of group and staff model HMO MCOs did not collect diagnostic code information.

Another survey conducted by the Health Research Center of the Park Nicollet Medical Foundation assessed the costs of implementing an ACG system. These costs provide a good sense of the costs of implementing a ADG-MDC or ADG-Hosdom system at each plan. This survey estimated that the start-up costs for a plan with 150,000 enrollees would be $712,500, and that yearly costs in the future would fall to $30,000 in 1994. Many of these costs are fixed, which would increase the unit costs of implementing and maintaining either risk adjustment system for plans with less than 150,000 enrollees.

There are numerous benefits to plans for collecting these diagnostic data. For example, plans may use these diagnostic data to improve the management of the care provided to their enrollees. In addition, these diagnostic data could also support provider profiling activities carried out by plans.

2. Creating a Capitated Risk Adjustment Payment System

The risk adjuster model estimates discussed throughout this paper and reproduced above in ***Table V-1*** are only the first step in creating a fully operational capitated risk adjustment payment system. Below we discuss several other parts of such a system. We caution that this list is by no means exhaustive, and we expect that all the steps needed to create a well functioning system will not be known until after preliminary risk adjustment payment systems are field tested in demonstrations and other trials.

The risk adjuster models developed for this project calculate unique payment scores for individuals in 1992 dollars. To use these models for payment purposes, HCFA will need to select some method of updating payment scores to the current payment year. In addition, the payment scores for the two JHU models are national payment scores. Given that the current AAPCC methodology HCFA now uses adjusts payments at the county level to reflect differences in the costs of providing care across different plan areas, it is also likely that some method of converting the JHU model's national payment scores to county payment amounts must also be designed.

One way of updating payment scores would be to use the percentage change in the U.S. per capita cost (USPCC) or some other factor to update these 1992 payment weights. There are at least two limitations of this approach. First, the relative payment weights are likely to change as medical practice and medical technology changes. Repeatedly updating an initial set of payment weights cannot account for these relative changes. Thus, HCFA may want to rebase its relative national payment weights by reestimating the regression equations underlying the ADG-MDC and ADG-Hosdom models using the most recently available data. In addition, the payment system may also be updated to reflect any ongoing developments in ADG, MDC, and Hosdom coding.

These two limitations could be solved in the following way. Suppose HCFA will be using either JHU risk adjuster model to set county-specific payment amounts in 1997. The first step would be to reestimate the model using individual demographic and diagnostic data from 1994 to estimate individual medical expenditures in 1995. These are the most recent years for which these data would be available for 1997 payment calculations. This would yield an updated set of national payment weights for 1995. These weights could then be adjusted for inflation to 1997, perhaps by using USPCC inflation rates.

In the second step, HCFA would compile 1996 demographic, diagnostic and cost data for each county. Using the cost data from 1996 for Medicare fee-for-service (FFS) enrollees for each county, HCFA would compute a county per capita cost (CPCC) amount.[3] These CPCCs would then be adjusted for inflation to convert them in 1997 amounts. Note that these CPCCs include no adjustments for any demographic or diagnostic factor. The CPCC in each county will indicate to plans the capitated payment amount they can expect to receive for an "average" enrollee in that county.

In the third step, HCFA would estimate a FFS average payment weight (FAPW) using one of the JHU models (that have been reestimated using 1994 and 1995 data) and the demographic and diagnostic data for FFS enrollees from 1996 in each county. HCFA would use the risk adjuster model to estimate a risk score for each FFS enrollee (this risk score would be equal to the sum of that individual's payment weights for each risk assessment factor). The county FAPW would then equal the sum of the risk scores for all FFS enrollees in that county divided by the total number of county FFS enrollees.

In the final step, HCFA would use the CPCCs and FAPWs to calculate payment amounts for risk contract enrollees in each county using the following formula:

$$\frac{Risk\ Score_{ij}}{FAPW_j} * CPCC_j$$

where i is an enrollee in a risk contract in county j. The payment amount for enrollee i thus equals enrollee i's risk score divided by the county FAPW and then multiplied by the county CPCC.

Depending on HCFA's objectives, the elements of this basic formula could change. For example, the APWs and CPCCs do not need to be county-specific, but instead could be either national averages or a blend of national and local averages. The first use of the JHU risk adjuster models by HCFA, however, will likely be in local demonstrations. In a local demonstration, the above formula would maintain budget neutrality at the county level.

3. Potential Perceptions of Risk Adjuster Methods by Plans

Risk adjuster models such as the two JHU models presented in this report represent a new and uproven technology. It will take time for MCOs and providers to become familiar with risk adjuster concepts, models and implementation. On the one hand, there appears to be considerable excitement among plans that improved risk adjuster models will allow the market for risk based contracts for Medicare patients to expand. With better risk adjustment, plans believe the downside financial risks of enrolling unusually high cost patients will be reduced.

In addition, some plans are eager to specialize in providing care to high cost patients. Plans believe the gains from providing care more efficiently to high cost patients are much

[3] HCFA may wish to make these calculations using moving average of several years of county FFS cost data, to smooth out any unusual yearly fluctuations.

greater than those of providing care to enrollees with fewer care needs. Plans have been reluctant in the past to enroll these patients, fearing that payments provided by the AAPCC do not adequately compensate them for the higher costs of heavier care patients. The use of more powerful risk adjuster models might counteract such behavior by plans and providers.

On the other hand, we are not sure plans appreciate some of the potential risks associated with risk adjusters. While plans are aware that better risk adjuster payment systems should increase their payments for high cost enrollees, they are less aware that their payments for healthier, lower cost patients are likely to decline significantly. Thus plans that now receive capitated payments under the AAPCC for younger, healthier Medicare enrollees are likely to receive smaller payments for these enrollees under a more sophisticated risk adjustment system.

These financial risks are compounded by the current marketing strategies of many Medicare MCO providers. Plans are now rewarded by attracting younger, lower cost enrollees, and the marketing strategies of plans reflect these incentives. If the relative payments for younger, lower cost employees decline under a diagnostic-based, risk adjusted capitated payment system, traditional Medicare MCO marketing strategies will need to change as well. Changing marketing strategies and identifying the most successful new strategies may be costly to plans. These uncertainties and costs for plans may reduce the enthusiasm plans and providers have for changing the current AAPCC payment system for Medicare risk contractors.

CHAPTER VI
DISCUSSION

This chapter re-assesses the major findings of the development and evaluation of the two JHU Medicare risk adjuster models. This chapter begins with a discussion of first the strengths and then the limitations of the ADG-MDC and ADG-Hosdom risk adjuster models. It then concludes with some suggestions for future research.

A. MODEL STRENGTHS

The potential strengths and weakness of the JHU models are discussed from the perspective of predictive accuracy; clinical and administrative cogency; data/logistical requirements; and potential strategic manipulation/gaming incentives.

1. Predictive Accuracy

Both JHU risk adjuster models are ***prospective*** – i.e., individual demographic and diagnostic characteristics in year one (1991) were used to predict medical expenditures in year two (1992). Predictive accuracy can be measured at either the individual or the group level. Accurate prediction at the individual level reduces the financial gains (losses) to plans that experience favorable (unfavorable) risk selection. This could encourage plans to compete on the basis of quality and efficiency rather than through risk selection. Predictive accuracy at the group level, particularly for high cost, non-random groups is also important. Risk adjuster models that accurately predict the expenditures of different groups of the population may improve the access to care of groups with predictable, high costs. Accurate prediction at the group level also helps ensure a plan's financial performance depends on that plan's efficiency and quality, rather than on the plan's mix of beneficiaries.

The individual predictive accuracy of the risk adjuster models developed during this project was measured using the adjusted R square statistic. Using this measure, the ADG-MDC model accounted for 6.0 percent and the ADG-Hosdom model accounted for 5.3 percent of the individual variance in year two medical expenditures.[1] This level of individual predictive accuracy for the two models is greater than it first appears. For example, each model represents a five to six-fold improvement in individual explanatory power as compared to a "AAPCC-like" model that explained one percent of the individual variation in year two medical expenditures. In addition, research has indicated that approximately 12 to 18 percent of the variation in individual medical expenditures in year two can be explained prospectively (Newhouse 89). Thus, the JHU models accounted for one-third to one-half of the individual variation in year two medical expenditures that could have been explained.

Predictive accuracy at the group level was measured using predictive ratios. For a given group, the predictive ratio is the ratio of predicted total medical expenditures to actual total medical expenditures for that group. Predictive ratios of less than 1.00 indicate under-prediction, while predictive ratios that exceed 1.00 indicate over-prediction. Exact prediction for a group occurs when that group's predictive ratio is equal to 1.00.

[1] The adjusted R-square statistics reported here were for the test split-half sample, which were slightly lower than those based on the development split-half sample.

Overall, the ADG-MDC and ADG-Hosdom models predicted the medical expenditures of groups more accurately than did the "AAPCC" model, as measured by the predictive ratio statistics. The predictive ratios for the two JHU models clustered more tightly around 1.00 than did those of the "AAPCC" model for repeated sets of random selectively groups of different sizes -- from 500 to 50,000. For non-random groups, especially for non-random groups defined by their level of medical expenditures or the presence of chronic diagnoses in 1991, the predictive ratios for the two JHU models were consistently much nearer to 1.00 than were the predictive ratios of the "AAPCC" model.

Both JHU models appeared to be equally good predictors for non-randomly selected groups, with one exception. The ADG-MDC model was a better predictor of expenditures for individuals with two or more hospitalizations. This was because the ADG-MDC model incorporated the number and type of previous admissions into the model.

2. Clinical Acceptance and Cogency

Clinical acceptance refers to the degree to which the methodology is intuitive to physicians and other clinicians and the degree to which it is based on medical and epidemiologic principles. This is an important criterion for at least two reasons. First, the diagnostic information upon which most risk adjusters are based derives from clinician input. Clinicians must be comfortable with a system's framework and find it understandable and clinically coherent. Second, given the organizational resources that are committed to capturing diagnostic information, it makes sense for a risk adjustment tool to serve double duty -- to support not only financing and payment activities, but also the many other clinically-oriented analytic activities now common in managed care organizations. These activities include the broad array of tasks that fall within quality improvement and utilization management.

A major strength of the ADG-MDC and ADG-Hosdom models was their clinical foundation. Both models were based on epidemiology and the natural history of disease, including the severity and likelihood of persistence of disorders. Over 100 HMOs and managed care organizations have found earlier versions of ADGs and ACGs to be useful for clinical profiling and related activities. Given this "real world" operational feedback, the clinical cogency and accuracy of the ADG classification system has been and will continue to be updated and improved. In addition, the DRG/MDC grouping of hospital diagnoses has already proven widely acceptable to clinicians as well as to non-clinician managers.

Even though the JHU models incorporated only selected ADGs and MDCs that were explicitly associated with future costs, the underlying ADG and MDC systems made use of all clinically related diagnoses. This allowed the diagnostic categorization systems, the building blocks of our models, to be used to for multiple purposes, such as describing the present and past epidemiologic and case-mix characteristics of a cohort.

3. Gaming

Any risk adjustment system can be gamed, usually in one of two ways. First, plans and providers may be better able to predict the future medical expenditures of their current or future enrollees than can a given risk adjuster model. If so, the plans can game the model by seeking enrollees who are favorable risks relative to that model. The current AAPCC model now used by HCFA may be susceptible to gaming through this type of biased selection (Hill et al., 1992). In contrast, both JHU models incorporate much of the data used by plans engaged in biased selection efforts, especially diagnostic information. If payments were based on either the ADG-MDC or ADG-Hosdom models, the gains from biased selection to plans and providers relative to the costs of selective enrollment and disenrollment will be sharply curtailed.

Second, plans and providers may manipulate information concerning enrollees, particularly diagnostic data, to "upcode" their enrollees into higher payment categories. While demographic risk adjustment models are virtually immune to plan and provider upcoding, diagnosis-based risk adjusters, including the ADG-MDC and ADG-Hosdom models, are not. It is not clear, however, whether the gains from upcoding would outweigh the costs of doing so were payments to be based on either JHU risk adjuster model.

4. Monitoring

HCFA may be able to reduce upcoding activities using unobtrusive, easily implemented and cost-effective monitoring activities, were HCFA to decide to use either JHU model as the basis of a risk adjusted capitated payment system. First, HCFA could identify the proportion of patients across time within a plan with "high cost" ADGs and MDCs or a hosdom diagnosis. These proportions should not change appreciably from year to year. Any sudden increases could indicate the possible presence of gaming.

A second monitoring activity would compare diagnosis patterns to treatments. This comparison would consist of random quality assessments of management strategies to determine whether patterns of diagnosis are consistent with implemented treatments. If a plan consistently provided fewer services than average to patients in a particular class, this would suggest that services were being skimped, or that their condition did not warrant the particular clinical designation. Some aspects of these assessments may be carried out using claims data. Other aspects may require medical record review and are more resource intensive. These activities could be integrated with other quality assurance and utilization review activities, such as those now conducted by federally qualified peer review organizations.

A third monitoring activity would be to require improvements in diagnostic coding. Specifically, diagnoses could be identified as "definitive" or "tentative." Currently, there is no way to distinguish definitive codes from those that are merely tentative (i.e., "rule-out codes). For example, patients in ambulatory settings often present undifferentiated symptoms and/or signs that do not permit the assignment of a definitive diagnosis. Conventional coding systems, such as the ICD-9-CM system, have no mechanism to record rule-out and/or probable diagnoses, and third party insurers often will not reimburse for care not coded as a "diagnosis." It is likely, however, that only definitive codes can be used to predict future medical expenses. Including an additional digit to indicate whether a code is definitive or tentative could substantially improve the performance of diagnosis based risk adjuster models, as well as improving the accuracy of these data for other applications. In addition, plans would not be rewarded for recording "rule-out" and/or probable codes and thus upcoding some patients, under a diagnosis-based risk adjustment system.

In the absence of including a new diagnostic nomenclature, it is possible that for conditions prone to rule-out coding, an ADG or hosdom marker could be assigned to a beneficiary only if the rule-out prone diagnosis code is present two or more times with an intervening period of, for example, 30 days. This would reduce the impact on payment predictability that rule-out codes introduce.

Even with the types of monitoring and diagnostic improvements described above, it is likely that through increased documentation of "true" diagnoses in the ambulatory setting, there will be an ADG "creep" across time. Although it will be necessary to reestablish payment rates periodically, the updated JHU models would utilize the increased documentation of ambulatory conditions and easily could reestablish a re-calibrated payment formula. This iterative process occurred as response to changes in hospital admission diagnoses after the introduction of DRG payments.

B. MODEL LIMITATIONS

Although the models developed here represent an advancement of the state-of-the-art of risk adjustment methodology, there are some additional limitations associated with these models. These limitation are due both to the intrinsic characteristics of the model, and to the scope of the simulation and critical assessment that Lewin-VHI was able to undertake as part of the evaluation phase. Given the ongoing nature of the research and development endeavor, some of these limitations are singled out as suggested areas for further research.

1. Discretionary Admissions

The JHU models do not explicitly include discretion variables. While the decision to hospitalize a patient with certain conditions may be more discretionary than others, the development team could not identify any ICD-9-CM codes for which admission was always discretionary. For certain patients, and under certain circumstances, it is appropriate to hospitalize a patient for every ICD-9-CM code. Unfortunately, claims data do not contain sufficient information to decide whether a medical decision for a particular patient was discretionary.

It is recognized that hospitalization can be a discretionary event, and that payment systems should not reward inappropriate admissions. Both risk adjuster models developed for this project seek to minimize discretionary admissions, as well as minimize intentional gaming. The ADG-Hosdom model attempts to address these concerns by including a variable that flags diagnoses that are likely, but not necessarily, treated in the hospital and by excluding explicit prior use variables. The ADG-MDC model may provide incentives to admit some beneficiaries to hospitals who can be cared for either in an inpatient or ambulatory setting. However, the relatively low additional year two payments that plans would receive for year one hospital admissions under the model, as well as the chance that beneficiaries with hospital admissions in year one may not re-enroll in year two, significantly limits these incentives for discretionary hospital admissions.

2. Noncontinuous Enrollees And Decedents

Partial-year enrollment represents a difficult practical issue that must be dealt with when implementing any risk adjustment system. The approach taken in this study was to focus on beneficiaries who were continuously enrolled for the first 12 months of the study period, including individuals who died during year two. The data for patients who died during year two were adjusted using the method employed by Boston University during their DCG development (described in ***Appendix II-4***). Limiting the analyses to these selected enrollees eliminated the complexity of accounting for factors such as mid-year enrollment and death during the "assignment" year.

JHU performed sensitivity analyses to learn the effect on the models' individual level predictive accuracy (adjusted r-square statistics) if individuals who were non-continuous enrollees during year one – i.e., those who turned age 65 during year one – were included. This analysis showed no real differences in terms of adjusted r-square statistics, p-values and size of coefficients in the models.

Moreover, within the Medicare program the elderly never lose Medicare coverage unless they die. Thus, while a beneficiary may switch from one contracting plan to another, the risk adjustment information will not be lost to the system as the individual will presumably enroll elsewhere within Medicare.

Another issue is the treatment of individuals who first become Medicare eligible at age 65. For most of these individuals, there will be no diagnostic or prior use data.[2] One option would be to assign them an interim capitated payment amount related to their demographic characteristics (e.g., age, gender, Medicaid eligibility, and prior disability).

C. SUGGESTED FURTHER RESEARCH

Following are three overall areas of potential research and development activities suggested for improving the ADG-MDC and ADG-Hosdom models.

1. Enhance the Hospital Dominant (Hosdom) Marker

Additional research in the project's risk adjuster methodology should include further development of the ADG-Hosdom model's hosdom variable. For example, the hosdom marker could be changed to require that more than one ambulatory diagnosis or an inpatient diagnosis be recorded before the hosdom marker is triggered. Another option would be to use ICD-9-CM classes, rather than specific four or five digit codes, to trigger the marker. Finally, the hosdom variable could be divided into several sub-categories where the payment weight could then vary across these sub-categories.[3]

2. Integrate Models With Reinsurance And Carve-Out Plans

Disease specific carve-outs currently are being proposed as a payment method within risk adjusted payment systems of some state health programs (such as New York, California and Kentucky), Medicaid "1115" waiver programs, and private sector plans. Typically, carve-outs have been used for very high cost individuals. Similarly, most capitated health plans utilize reinsurance at various dollar thresholds in order to protect their plan. Reinsurance is necessary because health plans, and risk adjusters, have difficulty predicting the most expensive (outlier) individuals.

It is possible to combine disease-based carve outs, reinsurance thresholds, and a diagnostic risk adjuster into a single, comprehensive payment system. As part of a separate contract study with HCFA-ORD, Lewin-VHI and JHU are mid-way in a project developing such a comprehensive, three-pronged risk adjusted payment system for the population younger than age 65. The diagnosis based risk adjuster of this three-part effort will likely be similar to the JHU models presented in this report.

Future Medicare research should attempt to further develop or replicate a similar three part, risk adjusted payment system that combines a risk adjuster model with high cost and/or disease specific patients excluded from the base capitation payment. To mimic reinsurance thresholds, JHU performed limited sensitivity analyses of the impact on the ADG-MDC and ADG-Hosdom models by truncating annual year two expenses at $50,000 and $100,000. This analysis indicated that if outliers are excluded in some manner from the base capitation rate, the overall accuracy of our models improve. Indeed, extending the work of the parallel Lewin-VHI/JHU "under-65" risk adjustment project to the Medicare population is a recommended area of further research and development.

---

[2] There may be diagnostic and prior use data for individuals who turn age 65 if they were previously eligible for Medicare (e.g., for those who were Medicare eligible due to their receipt of Social Security disability insurance benefits).

[3] During this project, we explored dividing the hosdom variable in this way, but the resulting models did not exhibit any improvements in explanatory power.

3. HCFA Demonstrations

Any new risk adjusted payment proposal will require multiple demonstrations before it can be implemented as part of a Medicare payment formula for HMOs or other capitated plans. Payment and delivery system demonstrations provide an excellent opportunity to test and evaluate the JHU models from several important perspectives – statistical, clinical and administrative. We suggest that initially both of the JHU models operate parallel to the AAPCC payment method, as well as to other alternative risk adjusters such as the DCGs. This approach of several new risk adjusters "shadow pricing" the AAPCC payments will allow a comprehensive and comparative critical assessment by administrators and clinicians at the health plan level, by HCFA analysts, researchers and actuaries, and by any other third party evaluators.

In conclusion, there is reason for some optimism that risk adjusted payments can be made powerful enough to support a level playing field in MCO competition for Medicare beneficiaries. In this new environment, competition based on premium price and quality, rather than through the selection of "good risks" or the avoidance of "bad risks" would then be encouraged.

REFERENCES

REFERENCES

American Academy of Actuaries. Risk Adjustment Work Group. Health Risk Assessment and Health Risk Adjustment – Crucial Elements in Effective Health Care Reform. May 1993.

Anderson G and Knickman J. Adverse selection under a voucher system: Grouping Medicare recipients by level of expenditure. Inquiry 1984; 21(2):135-143.

Anderson G and Knickman J. Patterns of expenditures among high utilizers of medical care services: The experience of Medicare beneficiaries from 1974 to 1977. Medical Care 1984; 22(2):143-149.

Anderson G, Steinberg EP, Holloway J and Cantor JC. Paying for HMO care: Issues and options in setting capitation rates. Milbank Memorial Fund Quarterly 1986; 64(4):1-15.

Anderson G, Lupu D, Powe N, Horn S, Antebi S, Whittle J and Steinberg E. Payment amounts for capitated systems. Report under contract number 17-C-98990/3. Johns Hopkins University. December 1989.

Anderson G, Steinberg EP, Powe NR, et al. Setting payment rates for capitated systems: A comparison of various alternatives. Inquiry Fall 1990; 27:225-233.

Ash A, Porell F, Gruenberg L, Sawitz E, and Beiser A. Adjusting Medicare capitation payments using prior hospitalization data. Health Care Financing Review Summer 1989; 10(4):17-29.

Beebe J, Lubitz J, and Eggers P. Using prior utilization information to determine payments for Medicare enrollees in HMOs. Health Care Financing Review 1985; 6(3):27-38.

Block B and Brennan JA. Reliability of morbidity data in a computerized medical record system. Proceedings of the AAMSI Congress. 1989; 21-30.

Brown R *et al.* Do Health Maintenance Organizations Work for Medicare? Health Care Financing Review 1993 15(1): 7-23.

Brown R in Luft H (ed). HMOs and the Elderly. Ann Arbor, MI: Health Administration Press. 1994.

Clark D, VonKorff M, Saunders K, Baluch W, Simon G. A chronic disease score with empirally derived weights. Medical Care 1995; 33:783-95.

Dick RS, Steen EB. The computer-based patient record: An essential technology for health care. Washington, DC: National Academy Press, 1991.

Doremus HD and Michenzi EM. Data quality: an illustration of its potential impact upon a Diagnosis-Related Group's case mix index and reimbursement. Medical Care 1983; 21:1001-1011.

Ellis RP and Ash A. The continuous-update diagnostic cost group model. Report under cooperative agreement 18-C-98526/1. Boston University. June 1989.

Ellis RP. A time dependent DCG model. Report under cooperative agreement 18-C-98526/1. Boston University. June 1990.

Epstein AM and Cumella EJ. Capitation payment: Using predictors of medical utilization to adjust rates. Health Care Financing Review Fall 1988; 10(1):51-69.

Fisher ES, Baron JA, Malenka DJ, Barrett J and Bubolz TA. Overcoming potential pitfalls in the use of Medicare data for epidemiologic research. American Journal of Public Health 1990; 80(12):1487-1490.

General Accounting Office. Health care reform: Considerations for risk adjustment under community rating. Report 94-173. Washington, DC: U.S. General Printing Office, 1994a.

General Accounting Office. Medicare: Changes to HMO rate setting method are needed to reduce program costs. Report 94-119. Washington, DC: U.S. General Printing Office, 1994b.

Greenwald L, Lubitz J, Beebe J, Lambert R. Risk selection and health risk pooling: the problem and proposal for reform. Health Care Financing Administration, Office of Research and Demonstrations, August 1992.

Health Care Financing Administration. "Study and Recommendations to Congress on Ways to Refine the Adjusted Average Per Capita Cost (AAPCC) and the Adjusted Community Rate (ACR)." 1988. Health Care Financing Administration. Washington, D.C.

Hornbrook M. (ed). Risk-based Contributions to Private Health Insurance. Advances in Health Economics and Health Services Research, Vol 12, 1991.

Hsia DC, Krushat WM, Fagan AB, Tebbutt JA, and Kusserow RP. Accuracy of diagnostic coding for Medicare patients under the prospective-payment system. New England Journal of Medicine 1988; 109:745-751.

Jencks SF. Accuracy in recorded diagnoses. JAMA 1992; 267:2238-2239.

Johnson RE, Hornbrook MC, Nichols GA. Replicating the chronic disease score (CDS) from automated pharmacy data. J Clin Epidemiol 1994; 47:1191-9.

Lubitz J, Beebe J and Riley G. Improving the Medicare HMO payment formula to deal with biased selection. Advances in Health Economics and Health Services Research 1985; 6:101-122.

Newhouse JP. Rate adjusters for Medicare under capitation. Health Care Financing Review Annual Supplement 1986; 45-55.

Newhouse JP, Manning WG, Keeler EB and Sloss EM. Adjusting capitation rates using objective health measures and prior utilization. Health Care Financing Review 1989; 10(3):41-54.

Physician Payment Review Commission. 1995 Annual Report to Congress. Washington, DC, 1995.

Revicki, DA. The dependability of medical encounter diagnostic information. Medical Care 1984; 22(7):661-669).

Robinson JC, Luft HS, Gardner LB and Morrison EM. A method for risk-adjusting employer contributions to competing health insurance plans. Inquiry 1991; 28:107-116.

Roos LL, Sharp SM and Wajda A. Assessing data quality: A computerized approach. Social Science Medicine 1989; 28(2):175-182.

Roos LL, Roos NP, Cageorge SM and Nicol JP. How good are the data? Reliability of one health care data bank. Medical Care 1992; 20(3):266-276.

Rossiter LF, Nelson LM and Adamache KW. Service use and costs for Medicare beneficiaries in risk-based HMOs and CMPs: Some interim results from the National Medicare Competition Evaluation. American Journal of Public Health 1988; 78(8).

Starfield B, Weiner JP, Mumford L and Steinwachs D. Ambulatory care groups: A categorization of diagnosis for research and management. Health Services Research 1991; 26(1):53-74.

Studney DR and Hakstian AR. A comparison of medical record with billing diagnostic information associated with ambulatory medical care. American Journal of Public Health 1981; 71:145-149.

The Jackson Hole Group. The 21st Century American Health Care System, Teton Village, WY, 1991.

The White House: Working Papers on Health Care Reform. Unpublished, Washington, D.C., February 1993.

Trapnell G, McKusick D and Genuardi J. An Evaluation of the Average Adjusted Per Capita Cost (AAPCC) Used in Reimbursing Risk-Basis HMOs Under Medicare. Actuarial Research Corporation, Falls Church, Virginia, April 1982.

Vladek B. The Medicare Choice Initiative. Proposed Solicitation. Baltimore: Health Care Financing Administration, 1995.

Weiner JP, Starfield B, Steinwachs D and Mumford L. Development and application of a population oriented measure of ambulatory care case-mix. Medical Care 1991; 29:452-472.

Weiner JP. Ambulatory case-mix methodologies: Applications to primary care research. In: Grady M. (edt) Primary Care Research: Theory & Practice U.S. Agency for Health Care Policy and Research, Rockville, MD, 1991. (Pub #91-0011)

Weiner JP, Powe N, Steinwachs D and Dent G. Applying insurance claims data to assess quality of care: A compilation of potential indicators. Quality Review Bulletin 1990; 16(12):424-438.

Weiner JP. Application of ACGs to Risk Adjustment. Paper presented at HCFA-ORD Conference on Risk Adjustment and Health Policy Reform. Baltimore, MD, September 1993.

CHAPTER I
APPENDICES

Appendix I-1
Original ADGs

ADG	Example ICD-9-CM
1 Time limited, minor	Dermatitis (692.9)
2 Time limited, minor, primary infections	Acute URI (465.9)
3 Time limited, major	Acute pericarditis (420)
4 Time limited, major, primary infections	Viral pneumonia (480)
5 Allergies	Allergic rhinitis (477.0)
6 Asthma	Asthma (493)
7 Likely to recur, discrete	Vaginitis (616.1)
8 Likely to recur, discrete, infections	Otitis media (382.9)
9 Likely to recur, progressive	Diabetic ketoacidosis (250.1)
10 Chronic medical, stable	Hypertension(401)
11 Chronic medical, unstable	Coronary atherosclerosis (414.0)
12 Chronic specialty, stable, orthopedic	Chondromalacia patellae (717.7)
13 Chronic specialty, stable, ENT	Hearing loss (389.9)
14 Chronic specialty, stable, eye	Refraction disorder (367.9)
15 Chronic specialty, stable, other	Polycystic ovaries (256.4)
16 Chronic specialty, unstable, orthopedic	Juvenile osteochondrosis (730.1)
17 Chronic specialty, unstable, ENT	Chronic sinusitis (473.9)
18 Chronic specialty, unstable, eye	Glaucoma (365.9)
19 Chronic specialty, unstable, other	Pseudotumor cerebri (348.2)
20 Dermatologic	Acne (706.1)
21 Injuries/adverse effects, minor	Ankle sprain (845.00)
22 Injuries/adverse effects, major	Tear of meniscus (836.0)
23 Psychiatric, time limited, minor	Acute reaction to stress (308)
24 Psychiatric, persistent or recurrent, stable	Panic disorder (300.01)
25 Psychiatric, persistent or recurrent, unstable	Schizophrenic disorders (295)
26 Signs/symptoms, minor	Headache (784.0)
27 Signs/symptoms, uncertain	Palpitation (785.1)
28 Signs/symptoms, major	Chest pain (786.5)
29 Discretionary	Sebaceous cyst (706.2)
30 See and reassure	Skin scar/fibrosis (709.2)
31 Prevention/administrative	Routine medical exam (V70.0)
32 Malignancy	Malignant neoplasm-breast (174)
33 Pregnancy	Pregnant state (V22.2)
34 Dental	Chronic gingivitis (523.1)

Appendix I-2
Original ACGs

Ambulatory Care Group (ACG) Categories	
ACG	ACG DESCRIPTION
1	Acute Minor, Age ≤ 1
2	Acute Minor, Age 2-5
3	Acute Minor, Age 6+
4	Acute: Major
5	Likely to Recur, without Allergies
6	Likely to Recur, with Allergies
7	Asthma
8	Chronic Medical, Unstable
9	Chronic Medical, Stable
10	Chronic Specialty
11	Ophthalmological/Dental
12	Chronic Specialty, Unstable
13	Psychosocial, without Psychosocial Unstable
14	Psychosocial, with Psychosocial Unstable, without Psychosocial Stable
15	Psychosocial, with Psychosocial Unstable, with Psychosocial Stable
16	Preventive/Administrative
17	Pregnancy
18	Acute Minor and Acute Major
19	Acute Minor and Likely to Recur, Age ≤ 1
20	Acute Minor and Likely to Recur, Age 2-5
21	Acute Minor and Likely to Recur, Age > 5, Without Allergy
22	Acute Minor and Likely to Recur, Age > 5, With Allergy
23	Acute Minor and Chronic Medical: Stable
24	Acute Minor and Eye/Dental
25	Acute Minor and Psychosocial Without Psychosocial Unstable
26	Acute Minor and Psychosocial With Psychosocial Unstable, without Psychosocial Stable
27	Acute Minor and Psychosocial with Psychosocial Unstable and Stable
28	Acute Major and Likely to Recur
29	Acute Minor/Acute Major/Likely to Recur, Age < 2
30	Acute Minor/Acute Major/Likely to Recur, Age 2-5
31	Acute Minor/Acute Major/Likely to Recur, Age 6-11
32	Acute Minor/Acute Major/Likely to Recur, Age > 5, Without Allergy
33	Acute Minor/Acute Major/Likely to Recur, Age > 5, With Allergy
34	Acute Minor/Likely to Recur/Eye & Dental
35	Acute Minor/Likely to Recur/Psychosocial
36	Acute Minor/Acute Major/Likely to Recur/Eye & Dental
37	Acute Minor/Acute Major/Likely to Recur/Psychosocial
38	2-3 Other ADG Combinations, Age < 17
39	2-3 Other ADG Combinations, Males Age 17-34
40	2-3 Other ADG Combinations, Females Age 17-34
41	2-3 Other ADG Combinations, Age > 34
42	4-5 Other ADG Combinations, Age < 17
43	4-5 Other ADG Combinations, Age 17-44
44	4-5 Other ADG Combinations, Age > 44
45	6-9 Other ADG Combinations, Age < 6
46	6-9 Other ADG Combinations, Age 6-16
47	6-9 Other ADG Combinations, Males Age 17-34
48	6-9 Other ADG Combinations, Females Age 17-34
49	6-9 Other ADG Combinations, Age > 34
50	10+ Other ADG Combinations
51	No Diagnosis or Only Unclassified Diagnosis
52	Non-Users

Appendix I-3
Original PACS Components

PAYMENT AMOUNT FOR CAPITATED SYSTEMS (PACS) COMPONENTS:

- Age
- Gender
- Disability status
- Chronicity level of admission (acute, acute w/ sequelae, chronic)
- Hospital admissions (0, 1, 2+)
- Use of Part B coverage (ie: expenses over the deductible amount)
- Major Diagnostic Categories:

MDC 01	NERVOUS SYSTEM
MDC 03 and 04	EARS/NOSE/THROAT/RESPIRATORY SYSTEMS
MDC 05	CIRCULATORY SYSTEM
MDC 06	DIGESTIVE SYSTEM
MDC 07	HEPATOBILLIARY SYSTEM/PANCREAS
MDC 08	MUSCULOSKELETAL SYSTEM/CONNECTIVE TISSUE
MDC 09	SKIN, SUBCUTANEOUS TISSUE & BREAST
MDC 10	ENDOCRINE/NUTRITIONAL/METABOLIC DISEASES
MDC 11	KIDNEY/URINARY TRACT
MDC 18	INFECTIOUS/PARASITIC DISEASES
MDC 19 and 20	MENTAL DISEASE/ALCOHOL/DRUG ABUSE
MDC 21	INJURIES/POISONINGS/BURNS
MDC 23	HEALTH STATUS FACTORS

CHAPTER II
APPENDICES

Appendix II-1
Assessment of Completeness of Ambulatory and Inpatient Diagnostic Codes for Use in JHU Risk Adjustment Model

This analysis assesses the degree to which the ICD-9-CM codes included in the ambulatory and inpatient Medicare claims files are adequate to derive clinically useful diagnostic formation for inclusion in our risk adjustment models.

This analysis assesses the degree to which all in-scope claims records, for both the JHU and Lewin-VHI half of the population (r = 1.24 million), are adequate to group into the Ambulatory Diagnostic Group (ADG) system. We ran all in-scope claims records for the ADG-Hosdom model. These included face-to-face Part-B (mainly ambulatory) HCFA 1500 forms and or ambulatory facility records. On these forms we included both the "line-item" ICD-9-CM and one or more "header form" diagnostic code. A diagnosis was considered missing if it was not present on a line-item and at least the first position on the header.

We also assessed the Part-A inpatient records, which had room for multiple diagnosis codes. We assessed all codes included in these records. The diagnosis was considered missing only if the primary diagnosis was not present.

The table that follows indicates the percentage of inpatient and Part-B claims diagnoses that were groupable using the ICD-9-CM to ADG "look-up table" developed by this project. This is presented for 1991 (the year with slightly higher non-match rates). An ICD-9-CM code did not match into an ADG category for one of three reasons:

- It was not present (i.e., the ICD-9-CM code was missing);
- It was "illegal" and not understood by our ADG look-up algorithm (The system does accept some "illegal" codes not conforming to ICD-9-CM standards. For example, if a fifth digit needed to complete a code is missing, in many cases the ADG system can still successfully categorize it); and
- It has not been considered by the grouping algorithm (because it is a rare condition).

Percentage of 1991 Claims Groupable into ADGs

	Groupable Percentage	Non-Groupable Percentage	Total Percentage
Ambulatory & Part-B Records N=6,753,958	97.4%	2.6%	100.0%
Inpatient Part-A N=263,952	96.1%	3.9%	100.0%
Total (All in-scope) N=3,017,910	97.4%	2.6%	100.0%

Source of claims - in-scope Part B and Part-A claims for entire study sample in 1991. See text for discussion of claims included in ADG categorization of ADG-Hosdom model.

Appendix II-2
Data File Development Plan

1. Select data from HCFA raw data files

a. Identify potential study population

The first step in the data preparation is to identify the potential study population. The study period is January 1991 through December 1992. The study population is defined as:

A 5 percent national sample, excluding initially the following beneficiaries:

1. In Public Health Service/Indian Health Service facilities
2. Originally or currently entitled due to ESRD status
3. With only Part A or only Part B coverage
4. Enrolled in HMOs for all or part of the study period
5. Who are Railroad retirees
6. Whose enrollment status is indeterminate due to missing enrollment file data

Selection of HIC numbers for people meeting the above conditions can be made directly through the use of HCFA's Decision Support Access Facility (DSAF) and the monthly denominator files. This task was be performed by JHU (and/or BU) personnel using HCFA computers. A file of HIC numbers was created and turned over to HCFA for creation of a cross-referenced finder file.

b. Create HCFA ID finder file

The ID numbers selected in the preceding step represent the current ID numbers. In order to select claims for people who have had a change in their ID number, it is necessary to obtain a list of cross-referenced ID numbers. This process involves the use of the complete enrollment database, which only HCFA personnel have access to. It therefore was HCFA's responsibility to provide a cross-reference list to JHU (and/or BU) based on the ID list that JHU created in the preceding step.

c. Exclude people with HIC number changes during the study period

Claims for a 5 percent sample population have already been selected and cleaned by HCFA (i.e., duplicates have been removed and credits and replacements have been applied). These existing claims files are collectively termed the 5 percent Sample Standard Analytical File (SAF). The potential study population is included in this 5 percent sample. However, the SAF is created based on terminal digit of the HIC number, and people who had a change to this terminal digit during the study period have an incomplete claims history in the SAF. Since it is not currently possible to access the 100 percent claims files in a timely fashion, we excluded anyone from the study population who had a change to the HIC number during the study period.

HCFA estimates that about 5.5 percent of the 5 percent sample have had a cross-reference number assigned at some point in time but not necessarily during the study period. In

order to determine if someone had a change to the HIC number during the study period, HISKEW files were be used from: 1) the start of the study period (or as close to the start as possible); 2) one year after the start of the study period; and 3) two years after the start of the study period. Each HISKEW record reflects the current ID and a single cross-reference ID. Therefore, by: 1) examining the initial HISKEW record for the presence of a cross-reference ID; and 2) excluding anyone who has a cross-reference ID on one of the subsequent HISKEW records that either does not match the original cross-reference ID or does not have a terminal digit that would place it in the 5 percent sample, it was possible to identify people who had a change to the HIC number and have an incomplete claims history in the SAF. These people were then be eliminated from the study population and the finder file.

It should be noted that the above method would not have worked if, within one year during the study period, someone was initially in the 5 percent sample, then had a change to the HIC number that resulted in them being out of the 5 percent sample, and then had second change to the HIC number placing them back in the 5 percent, and then had a third change to the HIC number also placing them in the 5 percent sample. In this case, the HISKEW IDs would follow the pattern:

Change	Current HIC #	Cross-reference #
0.	Yes, in 5% sample	N/A
1.	No	Yes (the last current HIC is recorded here)
2.	Yes	No
3.	Yes	Yes

In the above example, the original HISKEW record would indicate that the person was in the 5 percent sample, and the last HISKEW record (one year later) would also indicate that the person was in 5 percent sample, but in fact, at some time during the year, the person was not in the 5 percent sample and would have an incomplete claims history.

Since three changes to the HIC number in one year would be extremely rare, the approach should be effective in eliminating people with incomplete claims histories.

d. Select cleaned SAF claims

Selecting the clean SAF claims -- including physician, institutional part B, inpatient, home health, hospice and SNF claims -- can be performed directly by JHU (and/or BU) personnel using DSAF and the final finder file from the preceding step. The selected claims will probably be in standard variable-length format, although various other formats are possible.

2. Additional data adjustments

a. Adjust the study population

After preliminary analysis of the eligibility and claims data, further adjustment of the study population was needed. The adjustment process involved, for example, flagging and eliminating records from the eligibility files and corresponding claims from SAF files, or simply flagging records and eliminating them from analysis but not from the files.

b. Create procedure-based SAF extract records

The variable length SAF claims records were converted to procedure-based fixed length records, with administrative data removed. A list of data items included in these extract records is included as an appendix

c. Adjust expenditure fields

Exact methodologies for standardizing dollar amounts used to represent expenditures were chosen by JHU, BU and HCFA. In past work at JHU, this standardization involved updating existing RBRVS tables for missing values through the use of the claims data and creation of a DRG-based RVS table for impatient claims. The final adjustments were merged to the procedure-based extract records.

d. Assign ADGS and MDCs

The assignment of diagnostic grouping codes were performed using available and developed software. These codes were merged to the eligibility records.

3. Create person-based files

The analytic file creation involved producing person-based summary records that can be used more efficiently by statistical software than procedure-based records. The end result was person-based records that may be merged to the eligibility data.

Appendix II-3
Physician Payment Estimation Method

RBRVS Estimations

Allowed Charges

Allowed charges were used in cases where RVUs could not be used. For example, RBRVS STATUS = 'X' indicated that this service was excluded from the RBRVS system, if Geographic Practice Cost Index (GPCI) weights could not be found for a line item.

Modifiers

Any service provided by a physician in an inpatient setting with no modifier had the modifier recoded as '26'- professional component.

Where services have the modifiers below, RVU amounts were multiplied by the percents listed:

50 Bilateral procedure - 150%
51 Multiple procedures - 50%
54 Surgery and Pre-op care - 100%
55 Post-OP care - 100%
62 Two surgeons - 62%
80, 81, 82 or AS - for surgical services only - 10.6% for Physician assistants and nurse practitioners and 16% for physicians

Percentages based on HCPCS codes are as follows:

HCPCS	PERCENTAGE
10000-19499	19%
20000-29909	21
30000-32999	14
33010-37799	7
38100-38999	7
39000-39599	7
40490-49999	10
50010-53899	9
54000-55980	10
58000-58999	14
59000-59899	23
60000-60699	9
61000-64999	13
65092-68899	20
69000-69979	14

Bundled Services

Services with an RBRVS STATUS = 'B' had the RBRVS amount set to zero.

Services without GPCI or RVU values

This occurred quite frequently with the 1991 data because the carrier locality code (PLCLTY) was incorrectly recorded in about 30% of the line items. Additional reasons include incorrect HCPCS coding, and HCPCS codes not on the RVU table. The following logic was applied to these services:

1. Check to see if HCPCS is an anesthesia code (00100-01999). If so, then set RBRVS to allowed charges.
2. If the RVUs are not found, then first attempt to cross-walk 1991 codes. Next, try to standardize using median allowed charges for the code. Lastly, use allowed charges.
3. If the GPCI weights can be found, then attempt to use median allowed charges. If that fails, then use allowed charges.

Medians

In cases where the RVUs could not be calculated for a service, JHU had proposed substituting the median for all services with the same HCPCS code in that year. We lowered the threshold from the 95th percentile to the 50th percentile. In 1991 if a HCPCS code had at least $3,516.50 total allowed charges, or occurred 39 times, it could be standardized. Otherwise the RBRVS would be set to the allowed charges. In 1992 the cutoffs were $2,916.00 total allowed charges and 24 occurrences.

The standardized amount was put through the modifier logic explained above.

Number of Services

After the RBRVS amount is successfully calculated for the service, the value was multiplied by the (nonzero) number of services field. All other payment methods, such as using the allowed charges, were not multiplied.

Interventional Radiology

We used RVUs supplied by Nancy McCall, HER, Inc.

Person Level Summary File of Physician Expenditures

JHU keept the following variables in the person level summary file constructed from the Physician/SupplierSAF file. (All variables were aggregated separately for 1991 and 1992.)

1. Person ID number
   Variable: HICNO

2. Total Evaluation and Management visits
Variables: EMVIS91, EMVIS92

Sum the Number of Services field from line items with the following HCPCS codes:

90000-90199 OFFICE VISITS
90500-90599 ER SERVICES
90600-90699 CONSULTATIONS
90750-90774 PREVENTIVE MEDICINE
90801-90862 MENTAL HEALTH
92002-92140 OPHTHALMOLOGY
92502-92508 ENT SERVICES
95115-95199 ALLERGEN IMMUNOTHERAPY
98900-98922 CASE MANAGEMENT
99025-99025 MINOR SURGICAL PROCEDURE (PROFESSIONAL SERVICES)
99056-99062 SPECIAL SERVICES
99150-99152 PROLONGED DETENTION
59400-59430 OBSTETRICAL SERVICES
99201-99215 OFFICE OR OTHER O/P SERVICES (1992 revision)
99241-99245 OUTPATIENT CONSULTATIONS (1992 revision)
99281-99285 EMERGENCY DEPARTMENT SERVICES (1992 revision)
99341-99353 HOME CARE SERVICES (1992 revision)
99361-99373 CASE MANAGEMENT SERVICES (1992 revision)
99381-99397 PREVENTIVE MEDICINE (1992 revision)
99401-99429 COUNSELING AND/OR RISK FACTOR REDUCTION

3. Monthly and Yearly Totals (Payments, allowed charges and RBRVS amounts were summed into 24 monthly and 2 yearly totals.)

Variables:
Payments: PAY01 - PAY24, PAY91, PAY92
Allowed charges: ALW01 - ALW24, ALW91, ALW92
RBRVS: RVU01 - RVU24, RVU91, RVU92

Annual totals were be broken down as follows:

4. Physician Ambulatory Totals
Variables: PAPAYnn, PAALWnn, PARVUnn

Services in the range of the following HCPCS and accompanied by one of the following place of service codes:

99000 - 99499 1992 E & M
00100 - 01999 Anesthesia
10160 - 69979 Surgery (Excludes Maternal care)
77261 - 77799 Therapeutic radiology

78000 - 79999 Nuclear Medicine
90701 - 99199 Medicine (Includes 1991 E & M codes)

11 - Office
12 - Home
22 - Outpatient department
23 - Hospital emergency room
24 - Ambulatory surgical center
71 - State or local clinic
62 - Outpatient rehabilitation department
53 - Community mental health center
54 - Intermediate care facility for the mentally retarded

5. Physician Inpatient Totals
Variables: PIPAYnn, PIALWnn, PIRVUnn

Services in the range of the following HCPCS and accompanied by the following place of service code:

99000 - 99499 1992 E & M
00100 - 01999 Anesthesia
10160 - 69979 Surgery (Excludes Maternal care)
77261 - 77799 Therapeutic radiology
78000 - 79999 Nuclear Medicine
90701 - 99199 Medicine (Includes 1991 E & M codes)

21 - Hospital inpatient

6. Other Physician Service Totals
Variables: POPAYnn, POALWnn, PORVUnn

Services in the range of the following HCPCS and accompanied by one of the following place of service codes:

99000 - 99499 1992 E & M
00100 - 01999 Anesthesia
10160 - 69979 Surgery (includes maternal care)
77261 - 77799 Therapeutic radiology
78000 - 79999 Nuclear Medicine
90701 - 99199 Medicine (Includes 1991 E & M codes)

Place of service NOT equal to the places listed for physician ambulatory and inpatient totals.

Maternal care services were included in this category regardless of the place of service because they are billed differently than other physician services.

7. Physician Total
Variables: PTPAYnn, PTALWnn, PTRVUnn
The sum of a, b and c.

8. Ancillary Services
Variables: ANPAYnn, ANALWnn, ANRVUnn

Composed of two separate fields:
1. Lab - HCPCS Codes 80000-89399 (LAPAYnn, LAALWnn, LARVUnn)
2. X-ray/Imaging - HCPCS Codes 70010-76999 (XRPAYnn, XRALWnn, XRRVUnn)

9. All other line items
Variables: OTPAYnn, OTALWnn, OTRVUnn

Includes DME purchases, Levels II and III HCPCS codes, ext.

10. Patient copayment-adjusted Total RBRVS
Variables: DEDRVU91, DEDRVU92

The sum of all RBRVS amounts minus copayments and deductibles. This was computed as (Total RBRVS-300) * 0.80.

Appendix II-4
Decedent Adjustment Method

The JHU/Lewin VHI and the BU/CHER teams used the same method to account for the year-2 "partial enrollment" due to deaths (representing almost 5 percent of enrollees annually). The teams used weighted means and weighted least squares regressions, as described in an appendix by Ellis and Ash (Ellis and Ash 88a), and reiterated here with adaptations to the JHU models.

Adjustments in actual annual costs were needed in order for expected payments to equal expected costs for an entire group of enrollees. Conceptually, the adjustment process first expressed costs in annualized dollars. Second, when calculating means or conducting regressions, the observations were then weighted by the inverse of the weighting factor. Note that the resulting means and regression coefficients were not the same as the unweighted figures. This difference occured because the process of weighting and unweighting affected the numerator (the sum of annualized costs) and the denominator (the number of person-years) of the mean differently. This process of weighted least squares regressions and weighted means were widely used when each observation corresponds to a different sample size (in this case, number of months alive in year 2).

Numeric Formula

The weighting factor was calculated as using the inverse of the fraction of months that a person was alive in year 2. This reduced the weight attached to a person dying during the first few days of the year to twelve instead of up to 365. Also, since HMO payments were calculated as monthly payments, the use of months that a person was alive in an administratively relevant unit. Specifically, 1992 costs for each enrollee were converted to annualized costs by the following formula:

$$\text{Annualized costs} = \text{Actual costs}/W$$

where

$$W = \frac{\text{number of months enrollee was alive}}{12}$$

Regressions and means were all then calculated using annualized costs while weighting each observation by "W".

An Example with Two Enrollees

This section presents a hypothetical example with two enrollees to demonstrate how using weighted means generated the the correct payment. (An example with only two enrollees was chosen since all of the calculations can quickly and easily be verified visually.)

Suppose that there were only two enrollees, each of whom cost $1,800 in 1992. Person A lived the entire year (12 months) while person B lived only 6 months. Total costs for these two individuals is $3,600; they lived for a combined total of 1 1/2 years or 18 months; and a correct monthly payment would be $3,600/18=$200. The average annualized cost of these two people (i.e. the average cost per person-year) was $200 per month*12 months = $2,400. The second and third columns of the following table show how taking the average of either the actual cost or the annualized costs of each enrollee would have lead to the wrong result.

Mortality Adjustments in a Simple Example with Two Enrollees

	Months Alive (1)	Actual Costs (2)	Annualized Costs (3)	Weight Factor (4)	Weighted Costs (5) (3x4)
A	12	$1,800	$1,800	1.0	$1,800
B	6	$1,800	$3,600	0.5	$1,800
Total	18	$3,600	$5,400	1.5	$3,600
Mean	9	$1,800	$2,700	0.75	$2,400*

* The mean for column (5) was calculated as the total for Column (5) divided by the total for column (4). (This is how standard statistical packages would perform the calculation.)

As shown, taking the simple averaged of the actual costs led to payments which were too low, while taking the average of the annualized cost to payments that were too high. Using the weighted average annualized costs, for which the denominator will automatically be the sum of the weights, led to the correct payment.

A Regression Example

The numerical example verified why the weighting system worked correctly for calculating means. Establishing that the approach also works in a regression setting could be done analytically but is easier to see with a numerical example below. Consider a case in which there were ten enrollees, all of whom cost $2,400 per year. Suppose that all except one enrollee lived for the entire year; the one exception lived for only three months. Suppose that two enrollees, including the one that died, triggered the Medicaid eligibility variable in one of the JHU risk adjustors, and that the other eight triggered no risk adjustor variables.

The correct monthly payments for this group were readily be calculated by hand. Overall, total costs were $2,400*10 = $24,000, and the total person months 9*12 + 3 = 111. Average monthly payments for the entire group would have been $24,000/111 = $216.21, which was equivalent to $2,594.59 per person-year. Controlling for Medicaid eligibility, the average cost of those that triggered no risk adjuster variables was $2,400*8/(12*8) = $200 per month or $2,400 per year. The average cost of those in the Medicaid group was $2,400*2/(12+3) = $320 per month, or $3,840 per year. This was $1,440 higher than the average for the larger group.

When regressing annualized costs on only a constant term and without the Medicaid variable without any weighting, average payments were $3,120, which was too high. When

regressing annualized expenditures on a constant term (without the Medicaid variable) yet where each observation was weighted by the fraction of the year the enrollee was alive, average payments per person year were predicted to be $2,594.59 -- which was the correct annualized payment when all enrollees were paid at the same rate. Finally, when regressing on the Medicaid variable as well, the predicted payments for those without Medicaid were $2,400. The predicted payment for the Medicaid eligible were $1,440 higher. This confirms that the weighting technique also worked in a regression setting.

Regression on a Constant Without Weighting

Dependent variable: annualized expenditures

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	0	-1.11759E-05	5184	0.000	1.000
Error	9	46656000			
Root MSE		2276.84	R-Square	-0.0000	
Dep Mean		3121	Adj R-Sq	-0.0000	
C.V.		72.9756			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T For HO: Parameter=0	Prob>ITI
Intercept	1	3120.0	720.0	4.333	0.0019

Regression on a Constant With Weighting

Dependent variable: annualized expenditures

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	0	-7.45058E-09	0	0.000	1.000
Error	9	12509729.730	1401081		
C Total	9	12609729.730			
Root MSE		1183.673	R-Square	-0.0000	
Dep Mean		2594.595	Adj R-Sq	-0.0000	
C.V.		45.62072			

Parameter Estimates

	Parameter	Standard	T For HO:	Prob>ITI

Variable	DF	Estimate	Error	Parameter=0	
Intercept	1	2594.594	389.189	6.667	0.0001

Regression on a Constant and Variable with Weighting

Dependent variable: annualized expenditures

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	0	2241728.73	2241729.7	1.730	0.2249
Error	2	10366000.00	1296000.0		
C Total	9	12609729.73			
Root MSE		1138.423	R-Square	0.1778	
Dep Mean		2594.595	Adj R-Sq	0.0750	
C.V.		42.6766			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T For HO: Parameter=0	Prob>ITI
Intercept	1	2400.000	402.492	5.963	0.0003
Medicaid	1	1440.000	1094.89	1.215	0.2249

APPENDIX II-5
Percent of Variation in Subtotal Expenditures Explained

This appendix compares the adjusted R-square statistic when the dependent variable was total expenditures, inpatient hospital payments (physician and hospital fees), physician payments (inpatient and ambulatory service fees), or ambulatory payments (ambulatory physician service fees plus outpatient department fees).

PERCENT OF VARIATION IN SUBTOTAL 1992 EXPENDITURES EXPLAINED

MODEL	Total Payments	Hospital Payment	Physician Payment	Ambulatory Payment
ADG-MDC	.0625	.0563	.0581	.0718
ADG-HosDom	.0554	.0479	.0569	.0693
Demographic	.0102	.0097	.0049	.0017

Appendix II-6
Analytical Issues in Model Development

ADG system development

- Added several hundred new ICD-9-CM codes to ADG maps based on codes encountered among elderly in inpatient settings. The original ACG system used about 6,000 of the 10,000 ICD-9-CM codes. The original 6,000 codes accounted for over 95% of all encounters in the under-65 population.

- Recategorized ICD-9-CMs into existing ADGs based on clinical input and empirical analyses, using total payments as dependent variable.

- Modified several existing ADG categories to form new categories (e.g., new psychiatric ADG categories).

- Explored splitting the eye/dental ADG into major and minor categories. Currently minor conditions (e.g., myopia) are in with major conditions (e.g., glaucoma).

- Assessed impact of various ICD-9-CM code data sources on the explanatory power of ADGs (for example, ambulatory visit claims versus all claims).

- Assessed impact of modifications to the ACG algorithm. For example, 1) we added new a ACG with 15 or more ADGs, in order to take into consideration the higher proportion of sicker persons in the Medicare population. The original top ACG category, made up of 10 or more ADGs, was subdivided into ACGs comprised of patients with 10-14 ADGs and with 15 or more ADGs; 2) we split the "ADG combination" ACGs into ACGs based on "major" and "minor" ADG assignments; and 3) we explored different ACG age groups.

- Explored the impact of ADG total counts as an independent variable.

- Explored alternative methods to incorporate inpatient information, such as adding inpatient data to the ADG and ACG assignment source; adding prior use variables; developing a probable-admission variable.

- Developed a parsimonious ADG model where insignificant (p=.05) and negative ADGs are excluded from model.

PACS development

- Reviewed and updated ICD-9-CM codes, MDC assignments, and chronicity level assignments.

- Assessed impact of including partial levels of the chronicity assignment variable.

- Assessed impact of using various levels of admission count variable.

- Tested the use of MDCs as binomial variables instead of as count variables.
- Assessed impact of using age as a continuous variable, 10-year categorical variable, and 5-year categorical variable.
- Assessed impact of including a disability variable, and of looking at current disability status or past disability status.
- Recalibrated model with original "presence of Part B payments" variable excluded.
- Explored improving the predictive ability of certain MDCs by deleting selected DRGs from the MDC.
- Developed a parsimonious model where insignificant (p=.05) and negative MDCs are excluded from model.

Ambulatory/inpatient integration and development

- Developed "hospital dominant" variable. This variable is based on the presence of any number of 843 clinically and empirically selected ICD-9-CM codes. The variable is a binomial variable that reflects the presence of one or more diagnoses that are usually (over 50 percent of the time) treated in the hospital at least once during the year. Since this variable is not a measure of actual hospitalizations, it may be less prone to concerns regarding prior-use measures. We looked at 1) the number of occurrences of that diagnosis found in the inpatient claims of the JHU development data; 2) the number of occurrences of that diagnosis found among all claims of the JHU development data; and the ratio of the inpatient occurrences over total occurrences.
- Developed a model with ADGs and a "hospital dominant" binomial variable.
- Developed a model with ADGs and a variable where the 843 "hospital dominant" ICD-9-CM codes are grouped based on several ADG groupings.
- Considered using the hosdom variable as it is, except two or more ambulatory ICD-9-CM codes must be present, or one inpatient ICD-9-CM code must be present. (Currently, one ambulatory or one inpatient ICD-9-CM code triggers the hosdom variable.)
- Considered using the hosdom variable as it is, except the two ambulatory codes must be at least 30 days apart
- Tested an "ADG only" model, where ICD-9-CM codes from both inpatient and ambulatory claims are used to assign ADGs
- Assessed the conceptual and empirical compatibility between the PACS chronicity assignment, chronicity variable, and ADGs

- Assessed the interaction of ADGs and MDCs in a merged, ADG-PACS model
- Tested a basic ADG-PACS model, where ambulatory ICD-9-CM codes are used to assign ADGs and inpatient ICD-9-CM codes are reflected in the MDCs
- Tested numerous ADG-PACS model variants, where we included and excluded several combinations of original PACS variables with ADG assignments drawn from various claims files
- Developed a parsimonious, integrated model where insignificant (p=.05) and negative ADGs and MDCs are excluded from the model

Appendix II-7
Example of Several Rejected Models

A. sex, age

B. sex, age, ACGs

C. sex, age, ADGs

D. original PACS

E. PACS, ACGs

F. PACS, ADGs

G. sex, age, hosdom, ADGs

H. sex, age, hosdom, disabled, ADGs

I. sex, age, ADGs with only inpatient data, ADGs with only ambulatory data

J. sex, age, disabled, ADGs with only inpatient data, ADGs with only ambulatory data

K. sex, age, hosdom, disabled, chronicity, ADGs

L. sex, age, hosdom, disabled, count of admissions, ADGs

M. sex, age, hosdom, disabled, chronicity, count of admissions, ADGs

N. sex, age, hosdom, disabled, MDCs, ADGs

0. sex, age, hosdom, disabled, chronicity, MDCs, ADGs

Appendix II-8
Percent of Variation in Expendiutres Explained by Cell-based ACG Models

MODEL	ADJUSTED R-SQUARE
Age/Gender/ACGs	.0332
Age/Gender/ACGs/"Hospital Dominant" ACGs	.0428
Age/Gender/"minor" ACGs/ "major" ACGs	.0347
Age/Gender/ACGs where top ADG is split	.0336
Age/Gender/"minor" ACGs/ "major" ACGs and "top split" ADG	.0351

Appendix II-8
Percent of Variation in Expendiutres Explained by Cell-based ACG Models

MODEL	ADJUSTED R-SQUARE
Age/Gender/ACGs	.0332
Age/Gender/ACGs/"Hospital Dominant" ACGs	.0428
Age/Gender/"minor" ACGs/ "major" ACGs	.0347
Age/Gender/ACGs where top ADG is split	.0336
Age/Gender/"minor" ACGs/ "major" ACGs and "top split" ADG	.0351

CHAPTER III
APPENDICES

Appendix IV-1
ICD-9-CM Codes of Disease Based Population Cohorts for Predictive Ratios

Chapter IV states that the predictive ratios and adjusted R square statistics for several non-random groups are found in this Appendix. Described here are results for groups defined by the use of medical services in 1992 (***Table A-1***), and for groups defined by the level of their medical expenditures in 1992 (***Table A-2***).

These results are located in this Appendix rather than in Chapter Five because groups defined by year two service use are not optimal groups to use when evaluating risk adjuster models. Risk adjuster models use demographic, clinical, and prior use data in year one to predict medical expenditures in year two. The groups presented in ***Tables A-1*** and **A-2** are defined according to the actual use of medical services or actual level of medical expenditures in year two (1992).

None of the risk adjuster models is designed, or able, to predict the use of medical services or the level of medical expenditures by individuals in year two using year two service data. For example, the predictive ratios for the groups in ***Tables A-1*** and **A-2** are not close to 1.00 for any of the three models.

TABLE A-1: PREDICTIVE RATIOS AND ADJUSTED R SQUARE STATISTICS BY USE OF MEDICAL SERVICES IN 1992											
			AAPCC			ADG-MDC			ADG-Hosdom		
	N	Actual	Predicted	Adj. R. Square	Ratio	Predicted	Adj. R. Square	Ratio	Predicted	Adj. R. Square	Ratio
Number of Hospital Discharges in 1992											
None	489,002	$ 724	$ 3,430	0.26%	4.7381	$ 3,152	3.10%	4.3548	$ 3,156	3.06%	4.3599
One	86,034	$ 9,340	$ 3,751	0.13%	0.4016	$ 4,435	0.49%	0.4748	$ 4,440	0.61%	0.4754
Two	27,937	$ 18,917	$ 3,853	0.30%	0.2037	$ 5,182	0.73%	0.2740	$ 5,033	0.71%	0.2661
Three or More	17,534	$ 33,218	$ 3,906	0.74%	0.1176	$ 6,588	1.17%	0.1983	$ 5,095	1.21%	0.1534
High Users of Physician Services (1,527 RVUs or More in 1992) with No Hospital Admissions in 1992											
High Users	40,826	$ 10,367	$ 3,575	1.37%	0.3449	$ 4,698	5.35%	0.4531	$ 4,707	4.33%	0.4541

TABLE A-2: PREDICTIVE RATIOS AND ADJUSTED R SQUARE STATISTICS BY EXPENDITURE GROUPS IN 1992											
			AAPCC			ADG-MDC			ADG-Hosdom		
	N	Actual	Predicted	Adj. R. Square	Ratio	Predicted	Adj. R. Square	Ratio	Predicted	Adj. R. Square	Ratio
Expenditure Quintiles in 1992											
First	97,960	$ 0	$ 3,309	0.04%	106,160	$ 2,057	0.39%	65,998	$ 1,959	0.32%	62,857
Second	128,525	$ 116	$ 3,355	0.29%	28.88	$ 2,756	2.19%	23.73	$ 2,763	2.15%	23.79
Third	130,671	$ 455	$ 3,464	0.27%	7.6080	$ 3,388	1.34%	7.4422	$ 3,465	1.32%	7.6115
Fourth	131,283	$ 3,464	$ 3,554	0.30%	1.8202	$ 4,013	0.82%	2.0555	$ 4,115	0.56%	2.1078
Fifth	132,058	$ 14,817	$ 3,796	0.23%	0.2562	$ 4,967	1.63%	0.3359	$ 4,872	1.38%	0.3288
High Expenditures in 1992											
Over $50,000	7,411	$ 81,850	$ 3,865	0.16%	0.0474	$ 6,558	0.21%	0.0804	$ 5,972	0.35%	0.0732
Over $100,000	1,847	$163,602	$ 3,865	-0.10%	0.0236	$ 6,435	0.99%	0.0393	$ 6,052	0.92%	0.0370

Note: A "perfect" predictive ratio of 1.00 indicates that the risk adjuster model exactly predicted the expenditures of a given group. Predictive ratios of less than 1.00 indicate the risk adjuster model under-predicted the expenditures of a given group -- i.e., a predictive ratio of 0.90 indicates that the risk adjuster model under-predicted the expenditures of that group by 10 percent. Similarly, a predictive ratio above 1.00 for a risk adjuster model for a given group indicates that the risk adjuster model over-predicted the expenditures of that group.

```
'3004 ' = 'A depression       '
'3090 ' = 'A depression       '
'3091 ' = 'A depression       '
'311  ' = 'A depression       '


'303  ' = 'B alcohol and drug '
'3030 ' = 'B alcohol and drug '
'30300' = 'B alcohol and drug '
'30301' = 'B alcohol and drug '
'30302' = 'B alcohol and drug '
'30303' = 'B alcohol and drug '
'3039 ' = 'B alcohol and drug '
'30390' = 'B alcohol and drug '
'30391' = 'B alcohol and drug '
'30392' = 'B alcohol and drug '
'30393' = 'B alcohol and drug '
'304  ' = 'B alcohol and drug '
'3040 ' = 'B alcohol and drug '
'30400' = 'B alcohol and drug '
'30401' = 'B alcohol and drug '
'30402' = 'B alcohol and drug '
'30403' = 'B alcohol and drug '
'3041 ' = 'B alcohol and drug '
'30410' = 'B alcohol and drug '
'30411' = 'B alcohol and drug '
'30412' = 'B alcohol and drug '
'30413' = 'B alcohol and drug '
'3042 ' = 'B alcohol and drug '
'30420' = 'B alcohol and drug '
'30421' = 'B alcohol and drug '
'30422' = 'B alcohol and drug '
'30423' = 'B alcohol and drug '
'3043 ' = 'B alcohol and drug '
'30430' = 'B alcohol and drug '
'30431' = 'B alcohol and drug '
'30432' = 'B alcohol and drug '
'30433' = 'B alcohol and drug '
'3044 ' = 'B alcohol and drug '
'30440' = 'B alcohol and drug '
'30441' = 'B alcohol and drug '
'30442' = 'B alcohol and drug '
'30443' = 'B alcohol and drug '
'3045 ' = 'B alcohol and drug '
'30450' = 'B alcohol and drug '
```

```
'30451' = 'B alcohol and drug '
'30452' = 'B alcohol and drug '
'30453' = 'B alcohol and drug '
'3046 ' = 'B alcohol and drug '
'30460' = 'B alcohol and drug '
'30461' = 'B alcohol and drug '
'30462' = 'B alcohol and drug '
'30463' = 'B alcohol and drug '
'3047 ' = 'B alcohol and drug '
'30470' = 'B alcohol and drug '
'30471' = 'B alcohol and drug '
'30472' = 'B alcohol and drug '
'30473' = 'B alcohol and drug '
'3048 ' = 'B alcohol and drug '
'30480' = 'B alcohol and drug '
'30481' = 'B alcohol and drug '
'30482' = 'B alcohol and drug '
'30483' = 'B alcohol and drug '
'3049 ' = 'B alcohol and drug '
'30490' = 'B alcohol and drug '
'30491' = 'B alcohol and drug '
'30492' = 'B alcohol and drug '
'30493' = 'B alcohol and drug '


'4010 ' = 'C hypertension     '
'402  ' = 'C hypertension     '
'4020 ' = 'C hypertension     '
'40200' = 'C hypertension     '
'40201' = 'C hypertension     '
'4021 ' = 'C hypertension     '
'40210' = 'C hypertension     '
'40211' = 'C hypertension     '
'4029 ' = 'C hypertension     '
'40290' = 'C hypertension     '
'40291' = 'C hypertension     '
'403  ' = 'C hypertension     '
'4030 ' = 'C hypertension     '
'40300' = 'C hypertension     '
'40301' = 'C hypertension     '
'4031 ' = 'C hypertension     '
'40310' = 'C hypertension     '
'40311' = 'C hypertension     '
'4039 ' = 'C hypertension     '
'40390' = 'C hypertension     '
'40391' = 'C hypertension     '
'404  ' = 'C hypertension     '
'4040 ' = 'C hypertension     '
'40400' = 'C hypertension     '
'40401' = 'C hypertension     '
```

```
'40402' = 'C hypertension     '
'40403' = 'C hypertension     '
'4041 ' = 'C hypertension     '
'40410' = 'C hypertension     '
'40411' = 'C hypertension     '
'40412' = 'C hypertension     '
'40413' = 'C hypertension     '
'4049 ' = 'C hypertension     '
'40490' = 'C hypertension     '
'40491' = 'C hypertension     '
'40492' = 'C hypertension     '
'40493' = 'C hypertension     '


'401  ' = 'D hypertension
'4011 ' = 'D hypertension     '
'4019 ' = 'D hypertension     '


'2501 ' = 'E diabetes         '
'25010' = 'E diabetes         '
'25011' = 'E diabetes         '
'2502 ' = 'E diabetes         '
'25020' = 'E diabetes         '
'25021' = 'E diabetes         '
'2503 ' = 'E diabetes         '
'25030' = 'E diabetes         '
'25031' = 'E diabetes         '
'2504 ' = 'E diabetes         '
'25040' = 'E diabetes         '
'25041' = 'E diabetes         '
'2505 ' = 'E diabetes         '
'25050' = 'E diabetes         '
'25051' = 'E diabetes         '
'2506 ' = 'E diabetes         '
'25060' = 'E diabetes         '
'25061' = 'E diabetes         '
'2507 ' = 'E diabetes         '
'25070' = 'E diabetes         '
'25071' = 'E diabetes         '
'2508 ' = 'E diabetes         '
'25080' = 'E diabetes         '
'25081' = 'E diabetes         '
'2509 ' = 'E diabetes         '
'25090' = 'E diabetes         '
'25091' = 'E diabetes         '
'250  ' = 'F diabetes         '
'2500 ' = 'F diabetes         '
'25000' = 'F diabetes         '
'25001' = 'F diabetes         '
```

```
'428  ' = 'G cardiac          '
'4280 ' = 'G cardiac          '
'4281 ' = 'G cardiac          '
'4289 ' = 'G cardiac          '
'425  ' = 'G cardiac          '
'4250 ' = 'G cardiac          '
'4251 ' = 'G cardiac          '
'4252 ' = 'G cardiac          '
'4253 ' = 'G cardiac          '
'4254 ' = 'G cardiac          '
'4255 ' = 'G cardiac          '
'4257 ' = 'G cardiac          '
'4258 ' = 'G cardiac          '
'4259 ' = 'G cardiac          '


'410  ' = 'H cardiac          '
'4100 ' = 'H cardiac          '
'41000' = 'H cardiac          '
'41001' = 'H cardiac          '
'41002' = 'H cardiac          '
'4101 ' = 'H cardiac          '
'41010' = 'H cardiac          '
'41011' = 'H cardiac          '
'41012' = 'H cardiac          '
'4102 ' = 'H cardiac          '
'41020' = 'H cardiac          '
'41021' = 'H cardiac          '
'41022' = 'H cardiac          '
'4103 ' = 'H cardiac          '
'41030' = 'H cardiac          '
'41031' = 'H cardiac          '
'41032' = 'H cardiac          '
'4104 ' = 'H cardiac          '
'41040' = 'H cardiac          '
'41041' = 'H cardiac          '
'41042' = 'H cardiac          '
'4105 ' = 'H cardiac          '
'41050' = 'H cardiac          '
'41051' = 'H cardiac          '
'41052' = 'H cardiac          '
'4106 ' = 'H cardiac          '
'41060' = 'H cardiac          '
'41061' = 'H cardiac          '
'41062' = 'H cardiac          '
'4107 ' = 'H cardiac          '
'41070' = 'H cardiac          '
'41071' = 'H cardiac          '
```

```
'41072' = 'H cardiac          '
'4108 ' = 'H cardiac          '
'41080' = 'H cardiac          '
'41081' = 'H cardiac          '
'41082' = 'H cardiac          '
'4109 ' = 'H cardiac          '
'41090' = 'H cardiac          '
'41091' = 'H cardiac          '
'41092' = 'H cardiac          '


'411  ' = 'I cardiac          '
'4110 ' = 'I cardiac          '
'4111 ' = 'I cardiac          '
'1118 ' = 'I cardiac          '
'41181' = 'I cardiac          '
'41189' = 'I cardiac          '
'413  ' = 'I cardiac          '
'4130 ' = 'I cardiac          '
'4131 ' = 'I cardiac          '
'4139 ' = 'I cardiac          '
'414  ' = 'I cardiac          '
'4140 ' = 'I cardiac          '
'4141 ' = 'I cardiac          '
'41410' = 'I cardiac          '
'41411' = 'I cardiac          '
'41419' = 'I cardiac          '
'4148 ' = 'I cardiac          '
'4149 ' = 'I cardiac          '
'426  ' = 'I cardiac          '
'4260 ' = 'I cardiac          '
'4261 ' = 'I cardiac          '
'42610' = 'I cardiac          '
'42611' = 'I cardiac          '
'42612' = 'I cardiac          '
'42613' = 'I cardiac          '
'4262 ' = 'I cardiac          '
'4263 ' = 'I cardiac          '
'4264 ' = 'I cardiac          '
'4265 ' = 'I cardiac          '
'42650' = 'I cardiac          '
'42651' = 'I cardiac          '
'42652' = 'I cardiac          '
'42653' = 'I cardiac          '
'42654' = 'I cardiac          '
'4266 ' = 'I cardiac          '
'4267 ' = 'I cardiac          '
'4268 ' = 'I cardiac          '
'42681' = 'I cardiac          '
'42689' = 'I cardiac          '
```

```
'4269 ' = 'I cardiac          '
'427  ' = 'I cardiac          '
'4270 ' = 'I cardiac          '
'4271 ' = 'I cardiac          '
'4272 ' = 'I cardiac          '
'4273 ' = 'I cardiac          '
'42731' = 'I cardiac          '
'42732' = 'I cardiac          '
'42741' = 'I cardiac          '
'42742' = 'I cardiac          '
'4275 ' = 'I cardiac          '
'426  ' = 'I cardiac          '
'42760' = 'I cardiac          '
'42761' = 'I cardiac          '
'42769' = 'I cardiac          '
'4278 ' = 'I cardiac          '
'42781' = 'I cardiac          '
'42789' = 'I cardiac          '
'4279 ' = 'I cardiac          '


'490  ' = 'J pulmonary        '
'491  ' = 'J pulmonary        '
'4910 ' = 'J pulmonary        '
'4911 ' = 'J pulmonary        '
'4912 ' = 'J pulmonary        '
'49120' = 'J pulmonary        '
'49121' = 'J pulmonary        '
'4948 ' = 'J pulmonary        '
'4919 ' = 'J pulmonary        '
'492  ' = 'J pulmonary        '
'4920 ' = 'J pulmonary        '
'4928 ' = 'J pulmonary        '
'493  ' = 'J pulmonary        '
'4930 ' = 'J pulmonary        '
'49300' = 'J pulmonary        '
'49301' = 'J pulmonary        '
'4931 ' = 'J pulmonary        '
'49310' = 'J pulmonary        '
'49311' = 'J pulmonary        '
'4932 ' = 'J pulmonary        '
'49320' = 'J pulmonary        '
'49321' = 'J pulmonary        '
'4939 ' = 'J pulmonary        '
'49390' = 'J pulmonary        '
'49391' = 'J pulmonary        '
'494  ' = 'J pulmonary        '
'495  ' = 'J pulmonary        '
'4950 ' = 'J pulmonary        '
'4951 ' = 'J pulmonary        '
```

```
'4952 ' = 'J pulmonary          '
'4953 ' = 'J pulmonary          '
'4954 ' = 'J pulmonary          '
'4955 ' = 'J pulmonary          '
'4956 ' = 'J pulmonary          '
'4957 ' = 'J pulmonary          '
'4958 ' = 'J pulmonary          '
'4959 ' = 'J pulmonary          '
'496  ' = 'J pulmonary          '


'153  ' = 'K cancers            '
'1530 ' = 'K cancers            '
'1531 ' = 'K cancers            '
'1532 ' = 'K cancers            '
'1533 ' = 'K cancers            '
'1534 ' = 'K cancers            '
'1535 ' = 'K cancers            '
'1536 ' = 'K cancers            '
'1537 ' = 'K cancers            '
'1538 ' = 'K cancers            '
'1539 ' = 'K cancers            '
'154  ' = 'K cancers            '
'1540 ' = 'K cancers            '
'1541 ' = 'K cancers            '
'1542 ' = 'K cancers            '
'1543 ' = 'K cancers            '
'1548 ' = 'K cancers            '


'174  ' = 'L cancers            '
'1740 ' = 'L cancers            '
'1741 ' = 'L cancers            '
'1742 ' = 'L cancers            '
'1743 ' = 'L cancers            '
'1744 ' = 'L cancers            '
'1745 ' = 'L cancers            '
'1746 ' = 'L cancers            '
'1748 ' = 'L cancers            '
'1749 ' = 'L cancers            '


'162  ' = 'M cancers            '
'1620 ' = 'M cancers            '
'1622 ' = 'M cancers            '
'1623 ' = 'M cancers            '
'1624 ' = 'M cancers            '
'1625 ' = 'M cancers            '
'1628 ' = 'M cancers            '
'1629 ' = 'M cancers            '
```

```
'157  ' = 'M cancers          '
'1570 ' = 'M cancers          '
'1571 ' = 'M cancers          '
'1572 ' = 'M cancers          '
'1573 ' = 'M cancers          '
'1574 ' = 'M cancers          '
'1578 ' = 'M cancers          '
'1579 ' = 'M cancers          '


'433  ' = 'N stroke           '
'4330 ' = 'N stroke           '
'4331 ' = 'N stroke           '
'4332 ' = 'N stroke           '
'4333 ' = 'N stroke           '
'4338 ' = 'N stroke           '
'4339 ' = 'N stroke           '
'434  ' = 'N stroke           '
'4340 ' = 'N stroke           '
'4341 ' = 'N stroke           '
'4349 ' = 'N stroke           '
'436  ' = 'N stroke           '


'431  ' = 'O stroke           '


'820  ' = 'P hip fracture      '
'8200 ' = 'P hip fracture      '
'82000' = 'P hip fracture      '
'82001' = 'P hip fracture      '
'82002' = 'P hip fracture      '
'82003' = 'P hip fracture      '
'82009' = 'P hip fracture      '
'8201 ' = 'P hip fracture      '
'82010' = 'P hip fracture      '
'82011' = 'P hip fracture      '
'82012' = 'P hip fracture      '
'82013' = 'P hip fracture      '
'82019' = 'P hip fracture      '
'8202 ' = 'P hip fracture      '
'82020' = 'P hip fracture      '
'82021' = 'P hip fracture      '
'82022' = 'P hip fracture      '
'8203 ' = 'P hip fracture      '
'82030' = 'P hip fracture      '
'82031' = 'P hip fracture      '
'82032' = 'P hip fracture      '
'8208 ' = 'P hip fracture      '
'8209 ' = 'P hip fracture      '
```

```
'715  ' = 'Q arthritis        '
'7150 ' = 'Q arthritis        '
'71500' = 'Q arthritis        '
'71504' = 'Q arthritis        '
'71509' = 'Q arthritis        '
'7151 ' = 'Q arthritis        '
'71510' = 'Q arthritis        '
'71511' = 'Q arthritis        '
'71512' = 'Q arthritis        '
'71513' = 'Q arthritis        '
'71514' = 'Q arthritis        '
'71515' = 'Q arthritis        '
'71516' = 'Q arthritis        '
'71517' = 'Q arthritis        '
'71518' = 'Q arthritis        '
'7152 ' = 'Q arthritis        '
'71520' = 'Q arthritis        '
'71521' = 'Q arthritis        '
'71522' = 'Q arthritis        '
'71523' = 'Q arthritis        '
'71524' = 'Q arthritis        '
'71525' = 'Q arthritis        '
'71526' = 'Q arthritis        '
'71527' = 'Q arthritis        '
'71528' = 'Q arthritis        '
'7153 ' = 'Q arthritis        '
'71530' = 'Q arthritis        '
'71531' = 'Q arthritis        '
'71532' = 'Q arthritis        '
'71533' = 'Q arthritis        '
'71534' = 'Q arthritis        '
'71535' = 'Q arthritis        '
'71536' = 'Q arthritis        '
'71537' = 'Q arthritis        '
'71538' = 'Q arthritis        '
'7158 ' = 'Q arthritis        '
'71580' = 'Q arthritis        '
'71589' = 'Q arthritis        '
'7159 ' = 'Q arthritis        '
'71590' = 'Q arthritis        '
'71591' = 'Q arthritis        '
'71592' = 'Q arthritis        '
'71593' = 'Q arthritis        '
'71594' = 'Q arthritis        '
'71595' = 'Q arthritis        '
'71596' = 'Q arthritis        '
'71597' = 'Q arthritis        '
'71598' = 'Q arthritis        '
```

Appendix III-1
Truncated Results of Models

The three tables in this appendix present the payment weights for each prospective risk adjuster model estimated using three different dependent variables. The first dependent variable ("No Truncation") are the annualized 1992 medical expenditures of each individual, the basic version of each model presented in the text. Next, the second and third dependent variables ("Truncated at $100,000" and "Truncated at $50,000," respectively) truncate annualized medical expenditures in 1992 for each individual either at $100,000 or $50,000. That is, these versions of the risk adjuster models assume that individuals whose annualized medical expenditures are in excess of these thresholds have medical expenditures equal to the threshold amounts.

Truncated versions of risk adjuster models allow one to incorporate individual stop-loss reinsurance. Under stop-loss reinsurance, the prospectively set risk adjusted payments are meant to cover the expected costs of enrollees below the stop-loss threshold. Above that threshold, the primary insurer and the reinsurer share any additional medical expenditures of enrollees.

Payment Weights for "AAPCC" Risk Adjuster Model in Untruncated and Truncated Form			
Variable	**No Truncation**	**Truncated at $100,000**	**Truncated at $50,000**
Demographic Variables			
Base Expected Payment	$1,893	$1,870	$1,810
Male	733	700	639
Years Over Age 65	108	100	103
Ever Received Disability	1,895	1,835	1,729
Medicaid Eligible	1,316	1,269	1,183

Payment Weights for ADG-MDC Risk Adjuster Model in Untruncated and Truncated Form			
Variable	**No Truncation**	**Truncated at $100,000**	**Truncated at $50,000**
Demographic Variables			
Base Expected Payment	$608	$611	$608
Male	604	578	532
Years Over Age 65	67	66	66
Ever Received Disability	1,119	1,082	1,024
Medicaid Eligible	761	733	687
Visit ADGs			
Time Limited, Major (3)	542	530	510
Time Limited, Major Primary Infection (4)	734	704	652
Asthma (6)	818	817	783
Likely to Recur, Discrete (7)	225	231	247
Likely to Recur, Progressive (9)	965	903	832
Chronic Medical, Unstable (11)	1,345	1,315	1,257
Chronic Specialty, Unstable, Orthopedic (16)	604	684	698
Injuries/Adverse Effects, Major (22)	525	525	503
Psychiatric, Time Limited, Minor (23)	694	648	623
Psychiatric, Persistent or Recurrent, Major (25)	804	825	695
Signs/Symptoms, Uncertain (27)	460	453	440
Signs/Symptoms, Major (28)	551	646	535
Malignancy (32)	1,347	1,314	1,239
Major Diagnostic Categories (MDCs)			
Nervous System (1)	1,533	1,497	1,410
Ears, Nose, Throat, Respiratory System (3 or 4)	3,237	3,051	2,749
Circulatory System (5)	1,697	1,662	1,672
Digestive System (6)	1,759	1,706	1,564
Hepatobilliary System, Pancreas (7)	1,030	930	775
Musculoskeletal, Connective Tissue (8)	1,117	1,086	1,035
Skin, Subcutaneous Tissue and Breast (9)	1,762	1,749	1,646
Endocrine, Nutritional, Metabolic Systems (10)	2,938	2,781	2,482
Kidney, Urinary Tract (11)	2,526	2,449	2,131
Infectious, Parasitic Diseases (18)	3,061	2,816	2,375
Mental Disease, Alcohol, Drug Abuse (19 or 20)	1,957	1,939	1,866
Injuries, Poisonings, Burns (21)	1,882	1,833	1,827
Health Status Factors, Trauma (23 or 24)	1,481	1,422	1,318
Blood, Immunological, Myeloproliferative Diseases, HIV, AIDS (16, 17, or 25)	3,875	3,611	3,094
Transplants	3,944	3,665	3,096

Payment Weights for ADG-Hosdom Risk Adjuster Model in Untruncated and Truncated Form			
Variable	**No Truncation**	**Truncated at $100,000**	**Truncated at $50,000**
Demographic Variables			
Base Expected Payment	$434	$453	$471
Male	613	587	536
Years Over Age 65	64	63	63
Ever Received Disability	1,176	1,141	1,078
Medicaid Eligible	802	775	724
All ADGs			
Time Limited, Major (3)	663	629	584
Time L nited, Major Primary Infection (4)	1,503	1,400	1,252
Asthma (6)	1,216	1,194	1,119
Likely to Recur, Discrete (7)	365	359	357
Likely to Recur, Progressive (9)	1,696	1,569	1,404
Chronic Medical, Unstable (11)	1,415	1,390	1,333
Chronic Specialty, Unstable, Orthopedic (16)	593	644	673
Injuries/Adverse Effects, Major (22)	862	464	441
Psychiatric, Time Limited, Minor (23)	1,222	1,119	1,063
Psychiatric, Persistent or Recurrent, Major (25)	1,088	1,100	1,042
Signs/Symptoms, Uncertain (27)	568	552	529
Signs/Symptoms, Major (28)	753	741	703
Malignancy (32)	1,429	1,387	1,292
Hospital Dominant Marker (Hosdom)			
Probable Hospitalization Diagnosis	1,749	1,661	1,514

Appendix III-2
Retrospective Results of Models

We have prepared, retrospective versions of each model using 1991 data. That is, diagnosis-based risk assessors (the ADGs, MDCs, and the Hosdom variables) were coded using 1991 data to predict medical expenditures in that year for individuals within our sample. The retrospective versions of each model were estimated for three different dependent variables -- an untruncated dependent variable, and a dependent variable truncated at $50,000 and $100,000.

The retrospective versions of each model were the same as the prospective, year two (1992) models. That is, no attempt was made to use ADGs or MDCs that were excluded from the year two prospective model during our model development. Thus, it is possible that some risk assessors strongly associated with retrospective medical utilization were not included, because these assessors did not predict future medical utilization in the year two, prospective models.

Given these limits, the basic findings were still intriguing. For example, there was a substantial increase in individual predictive accuracy for the retrospective, year one models compared to their prospective, year two counterparts. For example, while the adjusted R square statistics for the prospective, year two models ranged from 0.0625 to 0.0886 for the ADG-MDC model and from 0.0554 to 0.0806 for the ADG-Hosdom models, this jumped to 0.6438 to 0.6919 for the ADG-MDC model and from 0.4085 to 0.4562 for the ADG-Hosdom models. The considerably higher adjusted R square statistics for the retrospective ADG-MDC model compared to the ADG-Hosdom model were probably due to the inclusion of the MDC variables. In a retrospective model, variables that indicate the number of hospital discharges in that year will be excellent predictors of medical expenditures.

APPENDIX III-2
Retrospective Results of Models
(1991 Data Predicting 1991 Expenditures)

Comparison Of Adj. R-Squares Across All Models

	"AAPCC"	ADG-MDC	ADG-HOSDOM
1992, Prospective			
Normal Payments	.0102	.0625	.0554
$100k Truncated	.0132	.0769	.0702
$50k Truncated	.0156	.0886	.0806
1991, Retrospective			
Normal Payments	.0124	.6438	.4085
$100k Truncated	.0129	.6624	.4227
$50k Truncated	.0142	.6919	.4562

JHU DATA, "AAPCC" DEPENDENT VARIABLE: Total Payments 1991					
Variable	Df	Parameter Estimate	Standard Error	T For H0: Parameter=0	Prob >]T]
Intercep	1	1597.046866	17.33285279	92.140	0.0001
Male	1	457.439588	17.51876281	26.111	0.0001
Yrsovr65	1	67.830392	1.22621998	55.317	0.0001
Everdisa	1	1389.514324	35.25638396	39.412	0.0001
Medicaid	1	1346.176584	29.17956131	46.134	0.0001

JHU DATA, "AAPCC" DEPENDENT VARIABLE: Total Payments 1991 Truncated at 100,000					
Variable	Df	Parameter Estimate	Standard Error	T For H0: Parameter=0	Prob >]T]
Intercep	1	1592.541094	16.94308254	93.994	0.0001
Male	1	452.880206	17.12481194	26.446	0.0001
Yrsovr65	1	67.831142	1.19864552	56.590	0.0001
Everdisa	1	1389.908171	34.46356067	40.330	0.0001
Medicaid	1	1336.7647 ;1	28.52338976	46.866	0.0001

JHU DATA, "AAPCC" DEPENDENT VARIABLE: Total Payments 1991 Truncated at 50,000					
Variable	Df	Parameter Estimate	Standard Error	T For H0: Parameter=0	Prob >]T]
Intercep	1	1563.156693	15.77889325	99.066	0.0001
Male	1	431.065457	15.94813570	27.029	0.0001
Yrsovr65	1	67.981654	1.11628446	60.900	0.0001
Everdisa	1	1340.809752	32.09550821	41.776	0.0001
Medicaid	1	1281.257020	26.56349700	48.234	0.0001

JHU DATA, ADG-MDC MODEL DEPENDENT VARIABLE: Total Payments 1991					
Variable	Df	Parameter Estimate	Standard Error	T For H0: Parameter=0	Prob >]T]
Intercep	1	120.749654	11.45666793	10.540	0.0001
Male	1	212.543084	10.61562552	20.022	0.0001
Yrsovr65	1	-10.091546	0.74877121	-13.477	0.0001
Everdisa	1	-76.168791	21.27729827	-3.580	0.0003
Medicaid	1	203.183920	17.63554348	11.521	0.0001
VADG9103	1	512.985123	14.85690698	34.528	0.0001
VADG9104	1	488.534289	19.12091043	25.550	0.0001
VADG9106	1	-32.441973	32.71289476	-0.992	0.3213
VADG9107	1	39.975262	12.64893664	3.160	0.0016
VADG9109	1	670.398212	21.80057251	30.751	0.0001
VADG9111	1	386.717839	11.23452061	34.422	0.0001
VADG9116	1	339.991830	31.89681818	10.659	0.0001
VADG9122	1	661.172269	17.56313662	37.645	0.0001
VADG9123	1	467.410055	48.12435155	9.713	0.0001
VADG9125	1	411.113098	32.28272814	12.735	0.0001
VADG9127	1	277.159289	13.28317495	20.865	0.0001
VADG9128	1	360.981057	12.09724214	29.840	0.0001
VADG9132	1	1409.945863	16.48976552	85.504	0.0001
MDC9101	1	6694.862924	31.64845054	211.538	0.0001
MDC9103	1	5600.147497	23.28087591	240.547	0.0001
MDC9105	1	6990.466492	14.56990117	479.788	0.0001
MDC9106	1	6397.489034	25.86829752	247.310	0.0001
MDC9107	1	6364.989001	49.56877102	128.407	0.0001
MDC9108	1	8383.171236	26.66687739	314.366	0.0001
MDC9109	1	5889.384392	54.17837585	108.704	0.0001
MDC9110	1	4751.528359	49.52006086	95.952	0.0001
MDC9111	1	5751.002858	42.01419655	136.882	0.0001
MDC9118	1	10429	69.36074083	150.356	0.0001
MDC9119	1	5294.172210	44.39760465	119.245	0.0001
MDC9121	1	6285.511299	99.76451160	63.003	0.0001
MDC9123	1	12205	66.27624818	184.148	0.0001
MDC9125	1	4742.064041	34.63806418	136.903	0.0001
MDC9126	1	18575	81.35382722	228.326	0.0001

JHU DATA, ADG-MDC MODEL DEPENDENT VARIABLE: Total Payments 1991 Truncated at $100,000					
Variable	Df	Parameter Estimate	Standard Error	T For H0: Parameter=0	Prob >]T]
Intercep	1	117.731860	10.90435760	10.797	0.0001
Male	1	211.926786	10.10386070	20.975	0.0001
Yrsovr65	1	-9.672577	0.71267397	-13.572	0.0001
Everdisa	1	-67.976413	20.25154876	-3.357	0.0008
Medicaid	1	200.485303	16.78535800	11.944	0.0001
VADG9103	1	519.867175	14.14067577	36.764	0.0001
VADG9104	1	488.547955	18.19911743	26.845	0.0001
VADG9106	1	-21.344762	31.13585074	-0.686	0.4930
VADG9107	1	42.948824	12.03914867	3.567	0.0004
VADG9109	1	664.906560	20.74959666	32.044	0.0001
VADG9111	1	391.343869	10.69291971	36.598	0.0001
VADG9116	1	350.680808	30.35911611	11.551	0.0001
VADG9122	1	653.725252	16.71644178	39.107	0.0001
VADG9123	1	436.999424	45.80434224	9.541	0.0001
VADG9125	1	409.116058	30.72642187	13.315	0.0001
VADG9127	1	277.993647	12.64281122	21.988	0.0001
VADG9128	1	365.497621	11.51405060	31.744	0.0001
VADG9132	1	1410.262012	15.69481643	89.855	0.0001
MDC9101	1	6683.062779	30.12272193	221.861	0.0001
MDC9103	1	5568.071798	22.15853665	251.283	0.0001
MDC9105	1	6946.812177	13.86750612	500.942	0.0001
MDC9106	1	6363.866297	24.62122220	258.471	0.0001
MDC9107	1	6366.430017	47.17912822	134.942	0.0001
MDC9108	1	8381.267629	25.38130362	330.214	0.0001
MDC9109	1	5877.890194	51.56651029	113.987	0.0001
MDC9110	1	4745.664537	47.13276631	100.687	0.0001
MDC9111	1	5734.091708	39.98874948	143.393	0.0001
MDC9118	1	10280	66.01695417	155.714	0.0001
MDC9119	1	5304.534798	42.25725672	125.530	0.0001
MDC9121	1	6230.807692	94.95500065	65.619	0.0001
MDC9123	1	12006	63.08116070	190.331	0.0001
MDC9125	1	4655.945576	32.96821037	141.225	0.0001
MDC9126	1	17446	77.43187023	225.312	0.0001

JHU DATA, ADG-MDC MODEL DEPENDENT VARIABLE: Total Payments 1991 Truncated at $50,000					
Variable	Df	Parameter Estimate	Standard Error	T For H0: Parameter=0	Prob >]T]
Intercep	1	100.929438	9.70872236	10.396	0.0001
Male	1	208.960647	8.99599792	23.228	0.0001
Yrsovr65	1	-6.864853	0.63453107	-10.819	0.0001
Everdisa	1	-63.752660	18.03101764	-3.536	0.0004
Medicaid	1	190.431700	14.94488593	12.742	0.0001
VADG9103	1	540.053695	12.59018642	42.895	0.0001
VADG9104	1	483.330667	16.20363022	29.829	0.0001
VADC9106	1	13.760663	27.72188344	0.496	0.6196
VADG9107	1	62.411090	10.71908646	5.822	0.0001
VADG9109	1	662.106326	18.47445585	35.839	0.0001
VADG9111	1	42[illegible]3267	9.52046810	44.180	0.0001
VADG9116	1	356.901635	27.03031579	13.204	0.0001
VADG9122	1	631.536394	14.88352620	42.432	0.0001
VADG9123	1	420.447918	40.78201192	10.310	0.0001
VADG9125	1	431.082468	27.35734741	15.757	0.0001
VADG9127	1	293.199382	11.25655893	26.047	0.0001
VADG9128	1	395.906986	10.25156407	38.619	0.0001
VADG9132	1	1421.130603	13.97391949	101.699	0.0001
MDC9101	1	6468.477534	26.81984164	241.183	0.0001
MDC9103	1	5359.986190	19.72890914	271.682	0.0001
MDC9105	1	6567.266873	12.34696914	531.893	0.0001
MDC9106	1	6163.730231	21.92156744	281.172	0.0001
MDC9107	1	6168.118290	42.00605610	146.839	0.0001
MDC9108	1	8261.896258	22.59830786	365.598	0.0001
MDC9109	1	5610.693450	45.91237282	122.204	0.0001
MDC9110	1	4583.204849	41.96477765	109.216	0.0001
MDC9111	1	5530.029106	35.60408421	155.146	0.0001
MDC9118	1	9178.027237	[illegible]7836207	156.146	0.0001
MDC9119	1	5207.353030	37.62385537	138.406	0.0001
MDC9121	1	585[illegible].474697	84.54342492	69.213	0.0001
MDC9123	1	10727	56.16447093	190.998	0.0001
MDC9125	1	4318.233000	29.35332947	147.112	0.0001
MDC9126	1	14072	68.94166144	204.121	0.0001

JHU DATA, ADG-HOSDOM MODEL DEPENDENT VARIABLE: Total Payments 1991					
Variable	Df	Parameter Estimate	Standard Error	T For H0: Parameter=0	Prob >]T]
Intercep	1	-578.504023	14.69303901	-39.373	0.0001
Male	1	277.408550	13.65553748	20.315	0.0001
Yrsovr65	1	-14.743522	0.96596630	-15.263	0.0001
Everdisa	1	131.241944	27.40723377	4.789	0.0001
Buyin91	1	291.420161	22.72023005	12.826	0.0001
Ahsdom91	1	6064.468018	22.59382938	268.413	0.0001
AADG9103	1	1907.166134	18.71058374	101.930	0.0001
AADG9104	1	2322.082694	22.91768518	101.323	0.0001
AADG9106	1	299.358052	40.27508928	7.433	0.0001
AADG9107	1	308.013456	16.07561767	19.160	0.0001
AADG9109	1	3230.938230	25.50989452	126.654	0.0001
AADG9111	1	772.790346	14.66426351	52.699	0.0001
AADG9116	1	1040.908245	39.48935682	26.359	0.0001
AADG9122	1	2514.659257	21.39588042	117.530	0.0001
AADG9123	1	2054.538419	56.43433767	36.406	0.0001
AADG9125	1	1797.110597	35.94544732	49.995	0.0001
AADG9127	1	628.483863	16.88463640	37.222	0.0001
AADG9128	1	1118.134215	15.76346036	70.932	0.0001
AADG9132	1	1345.749048	20.90045737	64.388	0.0001

JHU DATA, ADG-HOSDOM MODEL DEPENDENT VARIABLE: Total Payments 1991 Truncated at $100,000					
Variable	Df	Parameter Estimate	Standard Error	T For H0: Parameter=0	Prob >]T]
Intercep	1	-573.263842	14.19213116	-40.393	0.0001
Male	1	273.822762	13.18999962	20.760	0.0001
Yrsovr65	1	-14.259266	0.93303506	-15.283	0.0001
Everdisa	1	138.415530	26.47287986	5.229	0.0001
Buyin91	1	288.748391	21.94566316	13.157	0.0001
Ahsdom91	1	6054.522070	21.82357168	277.430	0.0001
AADG9103	1	1893.604542	18.07271174	104.777	0.0001
AADG9104	1	2287.995120	22.13638674	103.359	0.0001
AADG9106	1	310.259207	38.90205077	7.975	0.0001
AADG9107	1	307.071197	15.52757563	19.776	0.0001
AADG9109	1	3188.894745	24.64022376	129.418	0.0001
AADG9111	1	774.769277	14.16433666	54.699	0.0001
AADG9116	1	1051.675763	38.14310512	27.572	0.0001
AADG9122	1	2488.713195	20.60646260	120.423	0.0001
AADG9123	1	1988.911086	54.51040603	36.487	0.0001
AADG9125	1	1773.350975	34.72001284	51.076	0.0001
AADG9127	1	625.098859	16.30901369	38.328	0.0001
AADG9128	1	1118.445553	15.22606023	73.456	0.0001
AADG9132	1	1336.986646	20.18792927	66.227	0.0001

JHU DATA, ADG-HOSDOM MODEL DEPENDENT VARIABLE: Total Payments 1991 Truncated at $50,000					
Variable	Df	Parameter Estimate	Standard Error	T For H0: Parameter=0	Prob >]T]
Intercep	1	-546.295970	12.93639046	-42.558	0.0001
Male	1	256.903942	11.92999017	21.534	0.0001
Yrsovr65	1	-11.422573	0.84390443	-13.535	0.0001
Everdisa	1	128.380818	23.94398829	5.362	0.0001
Buyin91	1	269.990636	19.84924589	13.602	0.0001
Ahsdom91	1	5952.389429	19.73881753	301.558	0.0001
AADG9103	1	1812.271517	16.34626836	110.868	0.0001
AADG9104	1	2101.535122	20.02175010	104.963	0.0001
AADG9106	1	334.230103	35.18582993	9.499	0.0001
AADG9107	1	306.807546	14.04426308	21.846	0.0001
AADG9109	1	2956.622637	22.28640150	132.665	0.0001
AADG9111	1	788.397832	12.81125110	61.539	0.0001
AADG9116	1	1037.113453	34.49938455	30.062	0.0001
AADG9122	1	2338.132622	18.69224434	125.086	0.0001
AADG9123	1	1822.503128	49.30315594	36.965	0.0001
AADG9125	1	1717.821494	31.40329218	54.702	0.0001
AADG9127	1	611.811633	14.75105221	41.476	0.0001
AADG9128	1	1110.315161	13.77155073	80.624	0.0001
AADG9132	1	1320.199723	18.25942417	72.302	0.0001

CHAPTER IV
APPENDICES